EVALUATING HUMANOID EMBODIED CONVERSATIONAL AGENTS IN MOBILE GUIDE APPLICATIONS

A Dissertation
Presented to
The Academic Faculty

By

Ioannis Anastasiou Doumanis

In Partial Fulfilment
Of the Requirements for the Degree
Doctor of Philosophy in Computer Science

Middlesex University
September 2011

Acknowledgements

Thanks, first and foremost, to my supervisory team, Dr Ray Adams and Dr Serengul Smith, for supporting me over the years. I would like to thank specifically my first advisor Dr Ray Adams. I am grateful that Ray decided to take the advisor role back from the very beginning of my research work. Coming from a programming background, but with a genuine passion for HCI, it was hard to find a way to combine these two distinct areas into a single PhD thesis. Over the years, Ray encouraged me to pursue my ideas, but also to take into consideration the issues surrounding their implementation. I do not believe that I could have received better guidance to complete my work without Ray's help. Of course this work would not be possible without the support and guidance of many colleagues from the academic and the industrial sector. From the academic sector, I would like to thank Assistant Professor Mark Chavez, from the Nanyang Technological University and Mr Brent Rossen, from the University of Florida. Their help was invaluable in the implementation of the six prototypes. From the Industrial sector, I would like to thank Chant Product Support, Mr Dati Roberto from Loquendo, Mr François-Régis Chaumartin from Proxeme, Mr George Kountouris from Mellow Multimedia, Mr Tomas Estrada from Marketview, and finally associate Professor Ningzhong Liu from PartiTek. A number of individuals also helped tremendously throughout the years. Miss Maria Tzouros with the graphic design of the prototypes, Miss Fotini Singel with proof reading the Greek texts and Miss Arxountoula Tsakou who spent hours recording the short video clips in the castle of Monemvasia.

I thank my family for all the support over the years. I am especially thankful for my parents, Maria and Anastasios Doumanis, who supported me emotionally and financially towards the completion of this Ph.D program. The project was very demanding, and its completion would have been impossible without my family's financial support. Finally, my deepest gratitude goes to my dad Anastasios Doumanis, who deeply believed in the potential of the project from the beginning. You are the source of my courage and inspiration to complete the Ph.D study.

Table of Contents

Chapter 4: Related Experimental Work

Chapter 5: Research Methods for the Thesis

Chapter 6: Developing the Talos Toolkit

Chapter 7: ECA Visual Presence Studies

Chapter 8: ECA Features Studies

Appendix E:

List of Figures

ECA Visual Presence Studies: Experiment Three

ECA Visual Features Studies: Experiment Four

ECA Visual Features Studies: Experiment Five

ECA Visual Features Studies: Experiment Six

Appendix A

Appendix B

List of Tables

ECA Visual Presence Studies: Experiment One

ECA Visual Presence Studies: Experiment Two

ECA Visual Presence Studies: Experiment Three

ECA Visual Features Studies: Experiment Four

ECA Visual Features Studies: Experiment Five

ECA Visual Features Studies: Experiment Six

Appendix D

Appendix E

Thesis Abstract

Evolution in the area of mobile computing has been phenomenal in the last few years. The exploding increase in hardware power has enabled multimodal mobile interfaces to be developed. These interfaces differ from the traditional graphical user interface (GUI), in that they enable a more “natural” communication with mobile devices, through the use of multiple communication channels (e.g., multi-touch, speech recognition, etc.). As a result, a new generation of applications has emerged that provide human-like assistance in the user interface (e.g., the Siri conversational assistant (Siri Inc., visited 2010)). These conversational agents are currently designed to automate a number of tedious mobile tasks (e.g., to call a taxi), but the possible applications are endless. A domain of particular interest is that of Cultural Heritage, where conversational agents can act as personalized tour guides in, for example, archaeological attractions. The visitors to historical places have a diverse range of information needs. For example, casual visitors have different information needs from those with a deeper interest in an attraction (e.g., - holiday learners versus students). A personalized conversational agent can access a cultural heritage database, and effectively translate data into a natural language form that is adapted to the visitor’s personal needs and interests. The present research aims to investigate the information needs of a specific type of visitors, those for whom retention of cultural content is important (e.g., students of history, cultural experts, history hobbyists, educators, etc.). Embodying a conversational agent enables the agent to use additional modalities to communicate this content (e.g., through facial expressions, deictic gestures, etc.) to the user. Simulating the social norms that guide the real-world human-to-human interaction (e.g., adapting the story based on the reactions of the users), should at least theoretically optimize the cognitive accessibility of the content. Although a number of projects have attempted to build embodied conversational agents (ECAs) for cultural heritage, little is known about their impact on the users’ perceived cognitive accessibility of the cultural heritage content, and the usability of the interfaces they support. In particular, there is a general disagreement on the advantages of multimodal ECAs in terms of users’ task performance and satisfaction over non-anthropomorphised interfaces. Further, little is known about what features of an ECA

influence what aspects of the cognitive accessibility of the content and/or usability of the interface.

To address these questions I studied the user experiences with ECA interfaces in six user studies across three countries (Greece, UK and USA). To support these studies, I introduced: a) a conceptual framework based on well-established theoretical models of human cognition, and previous frameworks from the literature. The framework offers a holistic view of the design space of ECA systems b) a research technique for evaluating the cognitive accessibility of ECA-based information presentation systems that combine data from eye tracking and facial expression recognition. In addition, I designed a toolkit, from which I partially developed its natural language processing component, to facilitate rapid development of mobile guide applications using ECAs.

Results from these studies provide evidence that an ECA, capable of displaying some of the communication strategies (e.g., non-verbal behaviours to accompany linguistic information etc.) found in the real-world human guidance scenario, is not *affecting* and *effective* in enhancing the user's ability to retain cultural content. The findings from the first two studies, suggest than an ECA has no negative/positive impact on users experiencing content that is similar (but not the same) across different locations (see experiment one, in Chapter 7), and content of variable difficulty (see experiment two, in Chapter 7). However, my results also suggest that improving the degree of content personalization and the quality of the modalities used by the ECA can result in both *effective* and *affecting* human-ECA interactions. *Effectiveness* is the degree to which an ECA facilitates a user in accomplishing the navigation and information tasks. Similarly, *affecting* is the degree to which the ECA changes the quality of the user's experience while accomplishing the navigation and information tasks.

By adhering to the above rules, I gradually improved my designs and built ECAs that are *affecting*. In particular, I found that an ECA can *affect* the quality of the user's navigation experience (see experiment three in Chapter 7), as well as how a user experiences narrations of cultural value (see experiment five, in Chapter 8). In terms of navigation, I found sound evidence that the strongest impact of the ECAs non-verbal behaviours is on the ability of users to correctly disambiguate the navigation

instructions provided by a tour guide system. However, my ECAs failed to become *effective*, and to elicit enhanced navigation or retention performances.

Given the positive impact of ECAs on the disambiguation of navigation instructions, the lack of ECA-*effectiveness* in navigation could be attributed to the simulated mobile conditions. In a real outdoor environment, where users would have to actually walk around the castle, an ECA could have elicited better navigation performance, than a system without it. With regards to retention performance, my results suggest that a designer should not solely consider the impact of an ECA, but also the style and effectiveness of the question-answering (Q&A) with the ECA, and the type of user interacting with the ECA (see experiments four and six, in Chapter 8). I found that that there is a correlation between how many questions participants asked per location for a tour, and the information they retained after the completion of the tour. When participants were requested to ask the systems a specific number of questions per location, they could retain more information than when they were allowed to freely ask questions. However, the constrained style of interaction decreased their overall satisfaction with the systems. Therefore, when enhanced retention performance is needed, a designer should consider strategies that should direct users to ask a specific number of questions per location for a tour. On the other hand, when maintaining the positive levels of user experiences is the desired outcome of an interaction, users should be allowed to freely ask questions. Then, the effectiveness of the Q&A session is of importance to the success/failure of the user's interaction with the ECA. In a natural-language question-answering system, the system often fails to understand the user's question and, by default, it asks the user to rephrase again. A problem arises when the system fails to understand a question repeatedly. I found that a repetitive request to rephrase the same question annoys participants and affects their retention performance. Therefore, in order to ensure *effective* human-ECA Q&A, the repeat messages should be built in a way to allow users to figure out how to ask the system questions to avoid improper responses. Then, I found strong evidence that an ECA may be *effective* for some type of users, while for some others it may be not. I found that an ECA with an attention-grabbing mechanism (see experiment six, in Chapter 8), had an inverse effect on the retention performance of participants with different gender. In particular, it enhanced the retention performance of the male participants, while it degraded the retention performance of the female participants.

Finally, a series of tentative design recommendations for the design of both *affecting* and *effective* ECAs in mobile guide applications in derived from the work undertaken. These are aimed at ECA researchers and mobile guide designers.

Chapter 1 Introduction

Over the past few years, the world has seen tremendous progress in wireless technology and mobile devices, with wireless networks becoming more pervasive and providing more bandwidth, than ever before and mobile devices becoming progressively smaller and more compact. Although the latest generation of mobile devices (e.g., Apple iPad[1]), has brought significant improvements both in terms of hardware and interactive features, the user interface (UI) is still based on the graphical user interface (GUI) first introduced in desktop environments.

The improved hardware features have enabled designers to create cleaner UIs, but the new multi-touch style of interaction has forced them to ask for more compact information architectures (IA). The limited screen size and the size of the human fingertip make it difficult for users to swipe and touch screen icons, etc. This means that the GUI elements (and thereby the underlying IA) should support the completion of tasks in a limited number of actions. Without IAs that minimize the user's input and maximize the system output, the supporting GUIs become overwhelmingly complicated.

To address the above limitation, a number of mobile systems already offer information aggregation services. This means that the mobile system aggregates content from multiple resources on the World Wide Web (WWW) and pushes them to the user, formatted in a single presentation medium, with minimal user intervention (e.g., Siri[2] (Apple,2013) and its Android competitor Robin[3] (Magnifis,2013)). The output content is usually formatted in short natural language sentences that enable collaborative completion of tasks as easily as working with another human being. However, because the dynamic nature of the user's situation in mobile scenarios rapidly affects his/her ability to process, store and respond to information, there is a strong need for additional communication modalities in order to ensure user comprehension and understanding. Embodied conversational agents (ECAs), in addition to processing natural language, can provide multiple communication

[1] http://www.apple.com/ipad/
[2] http://www.apple.com/ios/siri/
[3] http://www.magnifis.com/wpress/

modalities for two-way human-device communication. These include speech recognition and generation, emotion recognition, use of body gestures and facial expressions to augment the information content. Such agents are already a reality for the latest generation of mobile devices (e.g., BlueMars for iPad[4]) and several more projects are on the way for more technological breakthroughs.

However, insufficient attention has been paid to the empirical evaluation of such interfaces. The direct consequence is that there is a near to absence of evidence on the potential impacts of ECAs on the users of mobile applications. In ECA research for stationary systems, some effects have already been established, but relating those effects to the user in mobile environments is yet to be done. Given this lack of knowledge, there is a potential risk associated with the introduction of ECAs into mobile devices. If the ECA does not actually enhance the service, or it is not appropriate for the particular situation, the user may perform poorly, become distracted and the entire interaction may collapse. The research presented in this thesis is an attempt to fill this gap and provide some empirical evidence on what the effects are on the user performing real tasks, with the help of an ECA under simulated mobile conditions.

1.1 Motivation

The starting point of this research was my interest in Embodied Conversational Agents (ECAs). The particular interest began during my European Master's dissertation at the University of Athens between September 2000 and October 2002. During that period, I had the opportunity to investigate the viability of ECAs in electronic Web retail applications. The positive results of this work encouraged me to seek out further domains for the application of ECA technology.

I was offered an opportunity to pursue this goal by the School of Engineering and Information Sciences at Middlesex University, when they accepted my application to study for a PhD. The school of EIS at Middlesex has an active research program in

[4] http:// bluemars.com/

several related fields, such as human-computer interaction, natural language processing, and computer graphics and animation.

Initially, I was intrigued by the possibility of continuing the work on ECAs in the electronic commerce domain and focus on a specific problem, such that of Natural Language Processing. However, soon I realized, that no matter the effort that is invested in the technical aspects of ECAs, the end-users are those who decide whether these interfaces are successful or not. Therefore, my focus shifted to the human aspects of interacting with such a life-like persona in electronic commerce applications. However, soon my research scope was refocused when I realized the need for such applications in the domain of cultural heritage. My interest was focused on Greece, one of the most culturally-rich and historical countries. The country has almost a near absence of info-structure to manipulate such content, which results in visitors experiencing very little of the country's historical and cultural background. In addition, such tools provide an additional source of revenue for the country as it is crowded every summer, with tourists from all over the world.

1.2 Aim and Objectives

The aims of this research are fourfold: 1) to provide insights on the impact of ECAs on the users of mobile tour guide applications, 2) to provide a set of design heuristics for the creation of effective ECAs, and mobile applications using ECAs, 3) to provide some understanding of the psychology of the users of such systems, and finally to 4) develop the tools needed to support the rapid prototyping of such applications.

To accomplish this aim, a number of objectives had to be met. Initially, a small-scale initial study was conducted in the field with a preliminary mobile guide system to investigate the feasibility of this research. The system enables the user to navigate a particular area and uncover information of interest about certain locations in that area, with the help of an ECA. The impact of this system to the user was evaluated via measures of both satisfaction and performance. The goal of the evaluation was to get some initial data on the users' requirements, identify some strengths and weaknesses of the system, and test the appropriateness of selected experimental methods and techniques. My experiences and results of this evaluation, led to the formulation of a

number of hypotheses about possible effects of the ECA on the user, the refinement of the experimental techniques, and the development of six prototype mobile tour guide systems.

The enhanced systems were designed with the goal of testing the following variables from our research framework: agent's visual presence, the agent's competence and its natural language and attention-grabbing abilities, task navigation complexity, and task information difficulty. All of these six systems offer advanced functionalities including: an ECA that uses synthesized speech and nonverbal behaviours to provide personalized tours of the castle, natural language processing abilities, recognition of physical locations, etc.

In order to evaluate the impact of an ECA on the user's experience of the prototype tour guide systems, six empirical studies were conducted. The first three, measured the impact of the presence of a multimodal ECA on the user's experience of the system. The final three, manipulated various features of the ECA (e.g., competence and attention grabbing abilities) and their impact on the user's experience of the system.

The first experiment, evaluated the impact of a multimodal ECA on the accessibility and usability of a prototype tour guide system. In this study, a prototype system with a multimodal ECA was compared with a non-ECA control. The potential effects of the ECA on information retention and navigation were both addressed.

The second experiment focused on information retention. In particular, the study evaluated the impact of an ECA on the retention of information under simulated mobile conditions. For this study, the information-enabled system was compared with a non-ECA control.

The third experiment focused on the problem of navigation and examined it separately from that of information processing. In particular, this study evaluated the impact of an ECA on the participant's navigational ability. For this study, the navigation-enabled variant was compared with a non-ECA control.

The other three experiments manipulated various ECA features as follows: The fourth study, examined the issues of building effective question-answering (Q&A) dialogues. In particular, it investigated the quality of the answers produced by two approaches to natural language processing: a script-based (Virtual People Factory, 2013) and a parser-based approach. For the parser-based approach, a search-and-match algorithm was developed, that was made open-source for the benefits of the ECA research community. Of particular interest, was how specific styles of questioning the systems would affect the users' retention performance and overall experience.

In the fifth study, I investigated the impact of ECA competence on the retention of information. It was assumed that competence in the particular scenario is based on the ability of the guide to effectively use non-verbal means, and pauses in speech to augment verbal communication about the various attractions of the castle. Therefore, an ECA featuring full body and face communicative behaviours was compared with another ECA without these behaviours.

The sixth experiment manipulated the attention-grabbing strategies of ECAs, and measured their impact on the user's retention performance. Two strategies for grabbing attention, one humorous and the other serious, were compared. In addition, in this experiment a novel method for evaluating the cognitive accessibility of ECA-based information presentation systems was validated. The method combines data from eye-tracking, facial expression and retention performance analysis.

In total, 91 participants took place in all six experiments from various backgrounds and cultures. I conducted experiments in three countries: Greece, UK and USA.

Another important objective of this work was to design an authoring toolkit to enable the rapid prototyping of systems. However, the focus of this work was not building a UI toolkit. Its core architecture is presented along with a series of heuristics that should guide the toolkit's UI design. The heuristics were produced based on a cognitive walkthrough of two existing toolkits.

The last and the most important objective that was met, was to analyse the results of the above studies, and present their findings as a list of design heuristics. This list should assist in the development of successful ECA, and mobile guide applications using ECAs, as well as provide some understanding of the psychology of the users of such systems.

1.3 Structure of the Thesis

The current thesis is divided into the following three areas: the literature survey, the design and implementation of experimental studies, and finally, significant contributions (and avenues for future work), and supplementary material.

The purpose of the literature review survey, presented in chapters 2 to 4, is to provide details of published research relevant to this study.

Chapter 2 gives a brief introduction on the domains of mobile guides and Embodied Conversational Agents. It also briefly presents the key developments in ECAs and ECAs designed specifically for mobile tour guide systems.

Chapter 3 reviews the theories and assumptions behind ECAs for both stationary and mobile computer systems.

Chapter 4 discusses the relevant empirical research in the area. In this chapter, the questions that the literature leaves unanswered are discussed, along with the questions that were explored in this research.

In the second part of the thesis, presented in Chapters 5 to 8, the user studies are described.

Chapter 5 reviews the relevant research methods and techniques that have been used to address the problems outlined in Chapter 4.

Chapter 6 provides an overview of the Talos toolkit, an open-source authoring toolkit to build and evaluate mobile guide applications with Embodied Conversational

Agents (ECAs). The core architecture of the Toolkit is presented along with a number of guidelines that should guide its UI design. Furthermore, the design, development and evaluation of the natural language component of the toolkit is presented and discussed.

In Chapter 7, the results of the first three experiments are presented. These experiments, are designed to explore the impact of the presence of a multimodal ECA on a number of aspects of the overall tour guide experience, namely information retention and navigation ability. The first experiment examined these issues from a more generic perspective, while the other two focused on more specific problems.

Chapter 8, details three additional experiments designed to explore the impact of different features of an ECA on the user's subjective experience and performance. The first study investigated the problem of natural language communication with an ECA, while the other two focused on certain attributes of the ECA's behaviour.

The conclusion of the thesis is in Chapter 9, which presents the contributions of this research to the ECA research community and highlights possible avenues for future research.

Chapter 2 Basic Concepts

Mobile Guide Systems and Embodied Conversational Agents (ECAs)

This chapter introduces mobile guide systems and embodied conversational agents. In the first part of the Chapter (§2.1), I provide a definition of what mobile guide systems are, and an outline of their limitations. In the second part of the Chapter (§2.2), I discuss the recent attempts to use Embodied Conversational Agents (ECAs), as a possible solution to these limitations. Moreover, I define what an embodied conversational agent is and what features it should have to match the requirements of the mobile guide domain. In the last section of the Chapter (§2.3), I introduce a number of characters in the history of embodied conversational agents and attempt to predict their future.

2.1 Mobile Guide Systems

Mobile guides, are systems that provide mobile users with local and location-based services (LBS), such as navigation assistance, where and when they need them most. The pace of progress in the area has been phenomenal. At the moment, there is a large number of research projects working on improving the user experience of such systems (e.g., by utilizing photographs of landmarks[5] to assist navigation (Hile *et al.* 2008)), and several commercial services available to pedestrians and car drivers (e.g., Google Maps Navigation[6] (Google, 2013)). The advent of the new generation of smartphones and ultra-light tablet devices, with integrated location and wireless communication technologies (e.g., GPS and WIFI) and more natural interfaces have made mobile guide systems accessible to more people and a necessity of the modern way of life.

However despite the impressive growth and popularity, mobile guide systems still suffer from a number of problems compared with a stationary system such as a

[5] Landmarks can be any remarkable physical object, which is situation either along the route (e.g., a statue) or distant from the route. In some cases, a unique part of the route (e.g., a house) can be the landmark itself.
[6] http://maps.google.co.uk

traditional desktop computer. The most important problem is the limitation of the available resources in the mobile scenario. These are problems with technological resources (e.g., limited display size and I/O options) as well as with user cognitive resources (i.e., the user's cognitive ability to process, store and respond to information). The latter type of problem is a more frequent phenomenon in the mobile environment, because of the constant change of the user's location and situation. For example, the user's position[7] and context are factors that can influence the performance and the way which the user interacts with the system: if s/he is riding a fast car or walking in a narrow alley, the navigational instructions that they receive will differ from those given to a pedestrian in an open area. Furthermore, the user's abilities and needs may have a strong impact not only on how s/he interacts with the system, but also on what services s/he requests. Users familiar with an area for instance, would probably prefer only aggregated explanations about known sites, whilst deaf-users would require information delivered in a manner most accessible to them.

In the last few years, there have been several attempts to create services and technologies that improve the user's experience of systems in this unique environment. Perhaps the most important advancement is the advent of multi-touch[8] interfaces that allow users to interact with a guide system in a more "natural" manner. This development, coupled with significant improvements in the way a user accesses the functionalities of the system (e.g., by fusing speech recognition and aggregated search interfaces), has alleviated some of the problems inherited from the "traditional" 2D-GUI interfaces of the older systems. The most well-known problem of direct manipulation interfaces is that, as the system becomes more powerful and sophisticated, the supporting interface typically becomes more complex as well. In mobile situations, this could limit the user's ability to control the system, but it could also prevent the user from initiating and completing tasks with it.

[7] There is no technology capable of measuring the current position precisely at all times. Electromagnetic devices like electronic compasses suffer from electromagnetic field interference, the Global Positioning System (GPS), does not work properly inside buildings or in narrow alleys, and light-based systems such as infrared beacons require a tight infrastructure.

[8] Multi-touch sensing allows the user to interact with a system with more than one finger at a time, as s/he would with manual operations in the everyday life.

Although the progress has been made is certainly in the right direction, there is still a significant way to go before interfaces of such systems become truly transparent and natural. There are already commercial services that use a new paradigm for interacting with such systems, that of a conversation (e.g., the Siri system (Apple, 2013)). A conversational interface is perhaps the most natural method of communicating with a system, as the user does not have to learn complex command structures and functionality to operate it (Lai *et al.* 2000). Some researchers (Cowell *et al.* 2003) have gone even further by suggesting augmenting human-system conversations with computer generated characters that use intonation, gaze patterns, gestures and facial expressions, in addition to words, for conveying information and affect (Massaro *et al.* 2000).

2.2 Embodied Conversational Agents

The term Embodied Conversational Agent (ECA) has been used to characterise embodied conversational interfaces in various application domains (e.g., the Real Estate Agent (REA) (Cassel *et al.* 1999)). Below, I define what an ECA is and I present the minimal features all ECAs should share. Further, I discuss what features an ECA should have to match the requirements of the domain of mobile guides and conclude with a working definition that it will be used throughout this thesis.

2.2.1 What's in a name?

In order to define what an embodied conversational agent (ECA) is I need to begin with the definition of an agent. An agent is *"a computer system that is situated in an environment and that is capable of autonomous action in this environment in order to meet its design objectives"* (Wooldridge 1999). An interface agent is an agent that visually appears on the interface with some form of graphical representation and within its domain is capable of autonomous actions, without requiring explicit directions from the user. These actions can be "intelligent", reactive, or even partially-reactive and scripted.

A simple example of an interface agent is the email sorting agent of Microsoft Outlook[9] (Microsoft, 2013). Once the user has created the necessary rules, the agent processes the incoming email messages and sorts them into different folders without having to be explicitly invoked by the user. Another example, of a more "intelligent" interface agent is the UCEgo (Chin 1991), that is designed to help users learn about the UNIX operating system using natural language. The UCEgo agent has a large knowledge database about how to use UNIX, and can proactively offer the user information about certain concepts and commands, correcting at the same time, any user misconceptions.

In recent years, a number of research fields have been introduced into the development pipeline of interface agents. An interface agent with integrated learning algorithms can "program itself", that is, acquire the knowledge it needs of the user and extend its knowledge databases accordingly. The research area of natural language understanding (NLU) and natural language generation (NLG) has made possible the realization of human-like dialogues with the user. Since human-to-human communication also involves gestures, facial expressions, and body language as well as locomotion, computer vision techniques and computer graphics techniques were included.

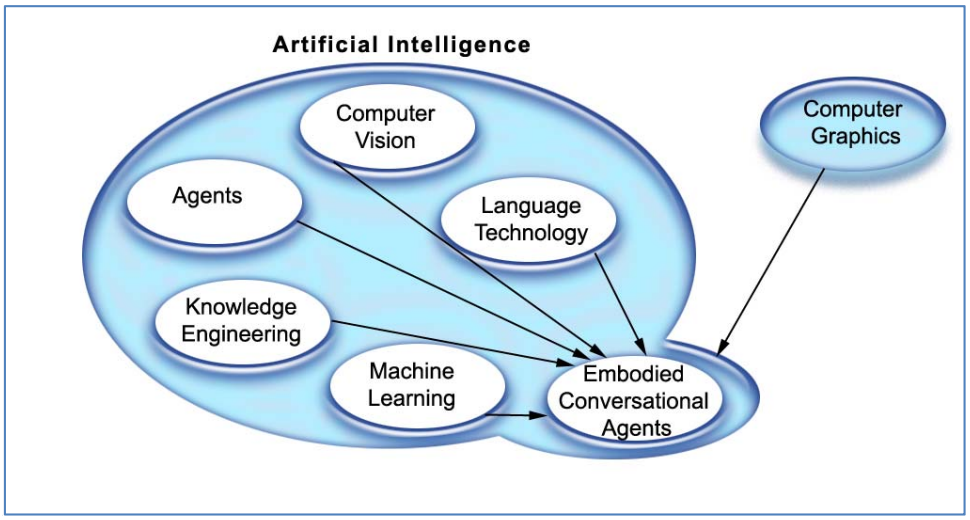


Figure 2.1: The origin of the ECA research area

[9] http://office.microsoft.com/en-gb/outlook/

Computer vision endows agents with the ability to recognize human gestures and facial expressions and react accordingly. Advanced computer graphic techniques enable the realization of more sophisticated visual agent representations, such as synthesis of gestures and facial expressions to accompany spoken natural language utterances. As a result of the combination of these different research fields, a new line of research was born, focusing on embodied conversational agents (see Figure 2.1).

Such agents visually appear on the screen as embodied characters, capable of engaging in conversations with the user and visualizing with their embodiment human conversational functions, such as gestures, head and body movements, etc. The way these components are elaborated varies: some sophisticated embodied conversational agents are fully equipped for making conversations, while in other embodied conversational agents a simple mechanism is used to process the input utterance of the user and produce an output through an animated character with synthesized speech and appropriate nonverbal behaviours. In both cases, however, embodied conversational agents are only capable of having conversations within a restricted domain. Hence it is questionable if this should be called a “conversation”. In some cases, it is better to speak of a human-computer interaction with elements of interpersonal communication. Despite this controversy, the term conversational serves well to describe the agents implemented for the purposes of this thesis work.

Other names used by researchers, such as; anthropomorphic agents, avatars, creatures, synthetic characters, life-like characters that are either, application-oriented (e.g., avatars for virtual reality environments) or highly subjective (e.g., life-like characters). For this reason, the working term that I am using throughout this thesis is the more widely accepted term embodied conversational agent (ECA). In the following section, I define the term embodied conversational agent, and extend it to match the requirements of the domain of mobile guides.

2.2.2 Minimal Requirements

In this section, I first attempt to define the features of existing embodied conversational agents, and then extend those features to match the requirements of the mobile guide field. I conclude with a working definition of an embodied

conversational agent that I use throughout this thesis. De Vos (2002) offers the following five features that all embodied conversational agents share:

1) Anthropomorphic Appearance

In the Oxford dictionary[10] anthropomorphism is defined as: "*the attribution of human characteristics or behaviour to a god, animal or object*". Hence, an anthropomorphic conversational agent is visually represented in the interface, by some form of anthropomorphic embodiment (human, animal or fantasy figure). This could be a real photograph, a 3D animation or any other visualization that can convey some human conversational functions.

2) Virtual body is used for communication purposes

The embodied conversational agent should be able to use its embodiment to communicate messages to the user or simply to enhance the on-going communication. This could be through facial expressions, gestures, body postures or animations that indicate the current state of the system.

3) Natural communication protocols

An embodied conversational agent uses very different communication protocols from Human-Computer Interaction (HCI) standards. These protocols draw from daily life human-to-human interactions; embodied conversational agents use natural language instead of buttons and menus.

4) Multimodality

An embodied conversational agent should be able to communicate through the natural modalities of human conversation. It must send and receive information through different communication channels that are typically used in face-to-face conversations, such as speech, gesturing and other modalities.

[10] Definition found at Oxford dictionary at, http://oxforddictionaries.com/definition/anthropomorphism

5) Social role

Embodied conversational agents, unlike normal computer programs, act out a social role copied from daily life in a believable manner. Often these roles are analogous to real world professions (e.g., tourist guides, teachers, etc.). Bates (1994) describes the notion of "believability" as the "*one that provides the illusion of life, and thus permits the audience's suspension of disbelief*". Research has shown two distinct pathways towards achieving "believability" in embodied conversational agents. The first pathway highlights the importance of introducing natural conversational functions into an ECA as a key element towards that direction. Some of the pioneers in the field, Cassel and her group (Cassel & Stone 1999) believe that implementing more natural language functions in an animated agent will result into higher "believability". The second pathway sees personality and emotions as an effective basis of a believable agent. Bates (1994) draws on the experience of Disney animators to support his argument that the portrayal of emotions plays a key role in the goal of creating "believable" characters. On a par with Bates, Trappl and Petta (1997) dedicated an entire volume to illustrate the value of the personality concept in ECA research. The work presented in this thesis, uses conversational agents that portray some emotions and personality. However, it does not examine whether such agents are more believable in their role as guides, in contrast for example, with a fully conversational embodied agent such as an agent controlled through a Wizard-of-Oz[11] technique. This question remains one of the open questions that the thesis poses for future research.

Based on the above requirements De Vos (2002) provides the following definition of an Embodied Conversational Agent (ECA):

"An ECA is an agent present in the interface, which mimics interpersonal communication when interacting with a user. To achieve this it uses human-like communication methods, appearance and tools."

[11] In a Wizard-of-Oz experiment, a subject interacts computer system that s/he believes to be autonomous but which is actually been operated or partially operated by an unseen human being with a

To fulfil the requirements of the mobile guide field, I add to the above list the following features:

6) Topological knowledge

An embodied conversational agent (ECA) should possess topological knowledge to be able to navigate the user within the physical environment (in real-time or at the user's request) and present information in a location-sensitive way. Moreover, the agent should be capable of exploiting this knowledge to answer the user's topological queries (e.g., how can I get from A to B) and location queries (e.g., what more can you tell me about this place?).

7) User-adaptable

The embodied conversational guide agent (ECA) should act as the user's personal guide, tailored to his/her needs, interests, personal and environmental context. For example, if the user of a tourist guide system is interested in architecture or history, the agent should be able to adapt the tour and the information provided by the system to meet these preferences.

In summary:

An embodied conversational agent is an anthropomorphic agent, running on the interface of a mobile guide system, which uses some (or the full range) of the human-like communication methods to interact with the user, is topologically-knowledgeable, user-adaptable, and is believable in its social role as a human guide.

This definition applies only to the ECAs in the domain of mobile guides, and not to ECAs in other domains. Therefore, whenever I refer to an ECA in the context of mobile guides, an ECA with the above definition applies. The use of the term ECA in any other context refers to an ECA as defined by De Vos (2002).

2.3 A Review of Embodied Conversational Agents

Embodied Conversational agents have been used successfully in a number of stationary and mobile guide applications. Recently, a number of projects have attempted to blend ECAs with the physical environment through a mobile augmented reality interface. Further, the current research has also attempted to look in the future of ECAs in mobile applications. These current and future applications are presented in detail below:

2.3.1 History

I start my review with one of the most recent developments in the mobile guide world, the Siri system (Apple, 2013). Siri is a mobile conversational agent (branded as a virtual personal assistant), that uses artificial intelligence and the user's personal context (e.g., location, time and preferences) to process natural language requests, along with service delegation to combine information from multiple internet resources and services, to generate an answer and return a combined output. The output is a mixture of natural language and additional web elements when it is required (e.g., a web form to book a place in a restaurant). Although Siri is not a full ECA, it has a lot of their characteristics.

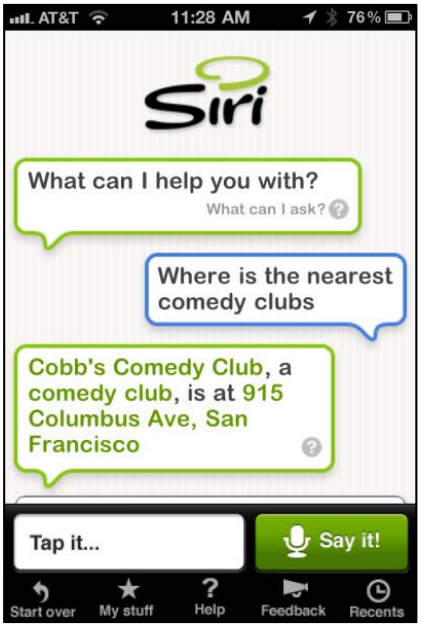


Figure 2.2: A user interacting with Siri (source: Google, 2010)

Firstly, it is fully conversational; it can process the intent of the user's request and generate the most appropriate responses using the collective knowledge of the World Wide Web. Then, it is well adapted to the user's context. For example, as is shown in the following example, the system uses the user's current location to disambiguate the word "nearest", and to return an appropriate answer that meets the request.

Similar systems to Siri, though not commercially available include: the GUIDE system (Cheverst *et al.* 2000), the LoL@ (Local Location Assistant) (Pospischil *et al*. 2002), the Deep Map System (Malaka & Zipf, 2000), the DELCA Ghost Project (DELCA 2004), and others. I believe that embodied conversational agents such as MACK (Stocky & Cassell 2002), and Ada and Grace (Swartout *et al*. 2010) from the realm of stationary information systems are the next step in the evolution of mobile ECAs. The research work presented in this thesis is a significant step towards this direction.

The MACK (Media Lab Autonomous Conversational Kiosk) (Stocky & Cassell 2002) agent, is a mixed - reality embodied conversational character capable of assisting users in finding a particular location in a building by means of direction giving (see Figure 2.3). The user can interact with the system using pointing gestures (on a device displaying the map of the building) along with natural language. The reactions of the agent are performed in a similar fashion, that is, it is capable of giving directions through speech and gestures to clarify the verbal instructions. These gestures may either happen on the screen of the kiosk (see figure 2.3 A, and figure 2.3 B), or in the real world by projecting the agent through a projector at the top of the map (see figure 2.3 C) to highlight a specific detail or region.

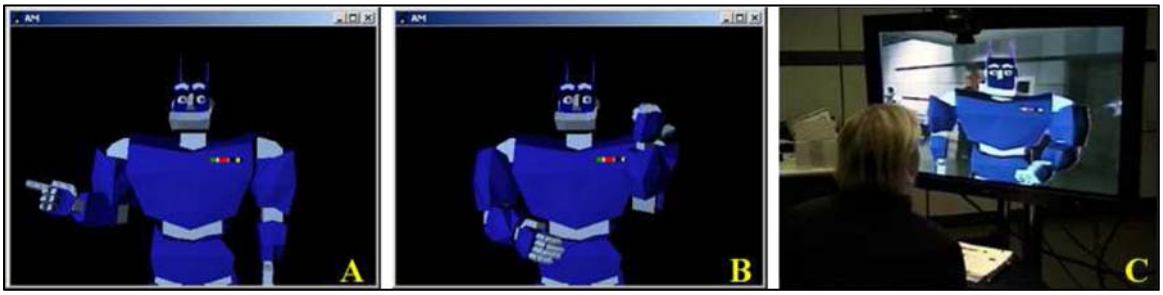
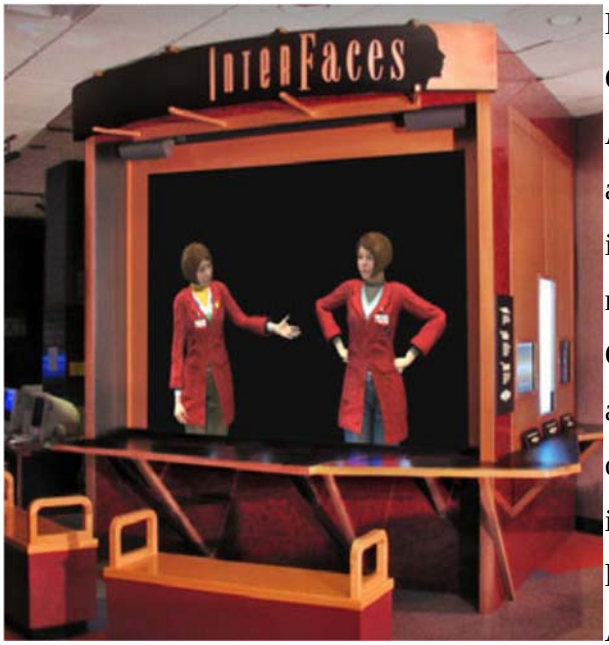


Figure 2.3: MACK in the information kiosk (screen A and screen B) and above the device displaying the map of the building (screen C) (source: Stocky 2002)

Ada and Grace (Swartout *et al*. 2010), are the twins embodied conversational guide agents (see Figure 2.4) of the Museum of Science in Boston, USA. The following example, demonstrates a typical scenario of a user interacting with the twins.



HANDLER: Why is this place named Cahners Computer Place?

ADA: Cahners Computer Place is named after Norman Cahners, a publisher based in Boston, and longtime supporter of the museum.

GRACE: Welcoming 300,000 visitors annually, Cahners Computer Place offers a one-stop resource for software that inspires people to create, explore and learn.

ADA: Did you read that in the brochure?!

The twins are also capable of responding to questions about their own exhibit and supporting technology:

HANDLER: What is your technology?

ADA: We're virtual humans. We use speech recognition technology to identify your words…

GRACE: [Finishing her twin's sentence] …and use statistics to figure out the meaning of the words based on context. Once we know what you're talking about, we'll reply appropriately.

Figure 2.4: A user interacting with Ada and Grace (source: Swartout *et al*. 2010)

The characters use a near photo-realistic appearance and natural language interactions accompanied by a full repertoire of human gestures to engage visitors with the museum contents. The combination of these elements created a *"jaw-dropping"* experience for visitors to the twins (Swartout *et al*. 2010).

In the realm of mobile applications, a limited number of projects have explored, the idea of embodied conversational agents. Some systems examine the problem of

dynamically generating believable story narrative/descriptions of attractions, while others focus more on human-agent conversations. At Heriot Watt University, Lim and Aylett (2007) built an affective guide system with an attitude that guides visitors in an outdoor attraction. Their guide dynamically generates story narratives about buildings on the University campus, with different emotional levels by taking into consideration various user and system factors (e.g., the degree of the user's interest to the stories, the guide's interests, etc.). The user gives feedback to the system solely through the GUI, and output is received in the form of speech, text and a simplistic 2D character. The animated character reflects through a range of facial expressions, the current emotional state of the system. In addition to storytelling, the system navigates the user to the chosen attractions via directional instructions.

The PEACH (Personalized Experiences with Active Cultural Heritage) (Kruger *et al*. 2007) prototype is a similar guide system designed for museums, featuring migrating characters. Such characters may be used in both mobile and stationary devices and may easily transit from one device to another. In one example, the characters could provide user and location adaptive multimedia presentations on the mobile device about the museum exhibits.

Figure 2.5: Virtual characters for mobile and stationary devices (source: Kruger *et al*. 2007)

In addition, when the information to be presented is not available in a format compatible with the mobile device, they could migrate to larger stationary systems spread throughout the museum, where they could deliver presentations with more sophisticated gestures and animations.

One of the few mobile conversational systems is the SmartKom mobile - the mobile device based version of the SmartKom System. In SmartKom (Malaka *et al.* 2004; Buhler *et al.* 2002) an anthropomorphic and affective user interface was realised in the form of an embodied character that combines speech, gestures and facial expressions for input and output. The mobile version of the system features a limited version of the character, offering route planning, and interactive navigation services through a city for pedestrians. Users interact with the system either through pointing gestures or speech utterances. The system outputs information in the form of synthesised speech, maps for displaying route information, and slide shows for augmenting any requested additional information about a point of interest nearby.

Recently, smart phones have gained access to several augmented reality applications (e.g., the Layar Browser[12] for the Android OS (Layar, 2013), and interest in the technology is increasing exponentially (e.g., Ellis, 2010). Hence, it is safe to say that Augmented Reality (AR) is potentially a very strong candidate platform for bringing mobile guide agent services and applications to the end-users. In addition, augmented reality interfaces are more “natural” to the nature of mobile guide agents, as humans are used to interact with other humans in the physical world, and not in some computer-generated environment. Below I review a number of early AR attempts.

The AR Puppet (Animated Agents in Augmented Reality) (Barakonyi & Schmalstieg, 2004) project utilizes an embodied guide agent to help users navigate inside a building. Figure 2.6, shows a screenshot of the system with the attributes that the agent receives, so it can deliver location-based information. These are: a) the selected destination of the user; b) the current and next room on the route suggested by the system; c) the next portal to go through; d) and a flag indicate whether the user has reached her/his destination.

[12] http://www.layar.com/

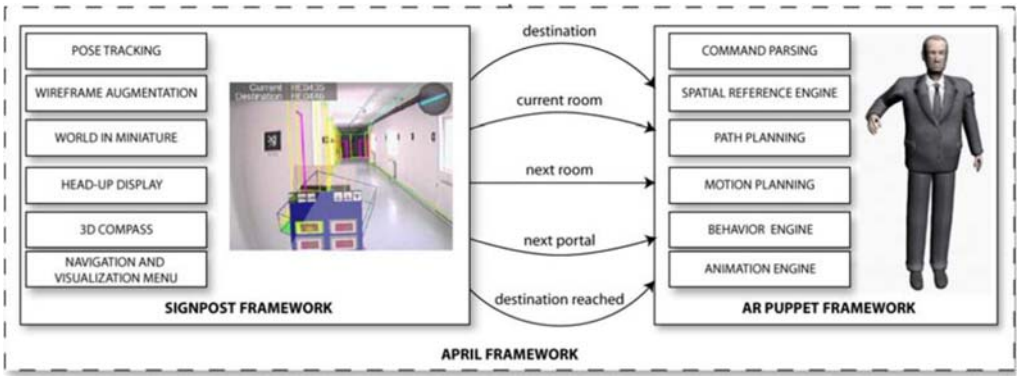


Figure 2.6: The Indoor AR navigation system interacting with a virtual tour guide (source: Barakonyi& Schmalstieg, 2004)

By watching these attributes the guide agent can deliver location-based descriptions and navigation instructions. Using appropriate gesturing (hand and head gestures), it is able to show the right direction, point out locations of interests in the building, inform when a door is being approached and notify when the destination has been reached.

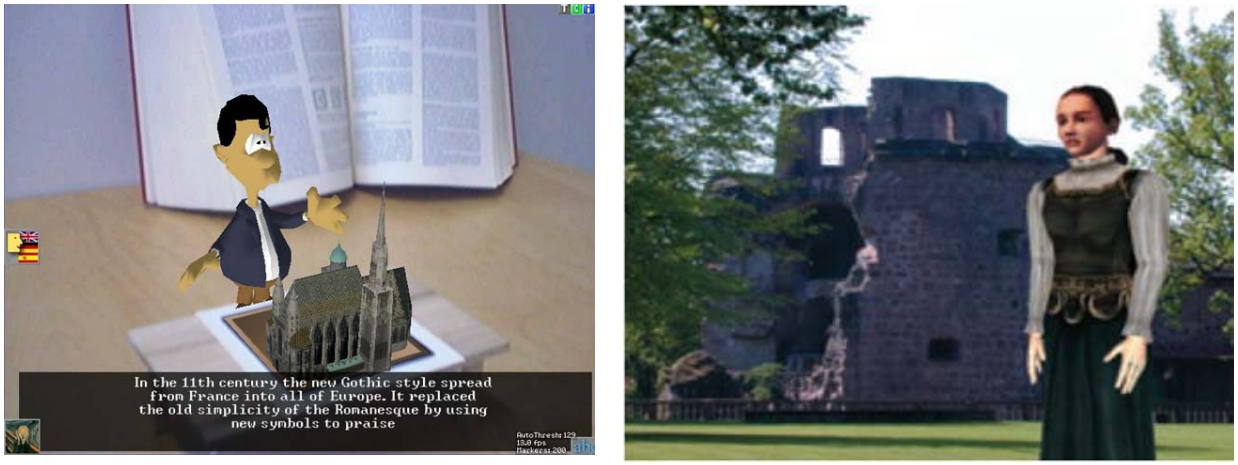


Figure 2.7: a) Mr Virtuoso an art history consultant in an AR game (source: Wagner & Schmalstieg (2006) b) a character of the GEIST system (source: Schneider *et al.* 2004)

Mr. Virtuoso (Wagner & Schmalstieg (2006 a) is a similar guide character acting as a consultant in art history in a collaborative educational game. The game requires players to sort artwork objects, according to their date of creation along a virtual timeline placed on the wall. If a player has problems, s/he can ask the expert advice of

Mr Virtuoso by placing the artwork on his desk. The character uses speech and text output to provide his advice, accompanied by a limited set of non-verbal behaviours.

GEIST (Braun 2003) is another innovative system using embodied guide agents in augmented reality. Within the GEIST system (see figure 2.7 b) , the history of the city of Heidelberg, Germany and the Thirty Years War is provided in a storytelling and a conversational manner to engage users in a dramatic, action-rich, and enjoyable experience. The storyline is revealed by virtual characters that can interact with the user and play the story as interactively altered by the user.

An AR platform that can significantly help (if it becomes commercially successful) to realise the full potential of mobile guide agents is the QderoPateo's Articulated Reality[13] (QPC, 2013) platform. Unlike most AR platforms that use 2D barcodes, the system aims for full image recognition. Figure 2.8 shows a potential application, where the user checks hotel room availability simply by pointing the device outside of the hotel.

Figure 2.8: QPC Augmented Reality application example (source: QPC, 2013)

The platform could be the basis for the first full-fledged embodied conversational agent in the market. Although detailed research and development is required, the agent could be fed data from the platform (e.g., from computer vision, accelerometers, GPS etc.) in order to create an internal knowledge model of its surroundings, the

[13] http://www.qpcmobility.com/cn/index.html

retrieved information, the user preferences and needs, and react accordingly rendering the output of the system fully multimodal.

From the discussion, it is very clear that interest in the field of mobile guide agents is growing. However current implemented systems use characters that are very superficial. In the current research, I aimed for a fuller user experience of mobile guide agents, similar to that of Ada and Grace (discussed above) in stationary systems. I have built applications for mobile devices, capable of full multimodal input and output (e.g., dialogues, human gestures to accompany the information provided, and others), I modelled interpersonal scenarios from real-world situations (see §7.2.2 of Chapter 7), and I extensively evaluated my systems under simulated mobile conditions.

2.3.2 Embodied Conversational Agents; the future

I believe that the notion of embodied conversational agents will inspire the design and development of several mobile applications and services. Besides the obvious applications in tourism (e.g., mobile or augmented reality electronic tour guide systems), another possible domain, is the development of sign-language guide agents to assist less-able people. Other applications include multilingual agents to assist people in everyday translation tasks (e.g., when they are in a foreign country), - and many more. As such systems learn to utilize more "intelligently" the resources of the World Wide Web (WWW) the more accurate and useful their output will become. An idea of mine is the creation of the "Global Guide Agent", a companion that provides access to information services, through multimodal methods of input and output, anytime, anywhere and regardless of the user's physical, cognitive, emotional and language background and the device s/he chooses to access the services. Currently my idea may sound farfetched, but the technology to create it is already here. Multi-core CPUs have recently become available on mobile devices, mobile graphic chipsets can handle complex 3D graphics, and the fourth generation (4G) of cellular networks promise to provide enough bandwidth to support any type of multimedia application.

2.4 Conclusion

Embodied Conversational Agents (ECA) have been introduced in this chapter, along with a selection of stationary and mobile systems that feature them. From this survey, it is clear, that many research projects have already explored the technical challenges involved in the realization of such systems. With regard to embodied guide agents, current research seems to focus on the tourist domain, where they are used to provide interactive information services to tourists during a visit. However the enormous effort that has been spent on improving the various areas of embodied guide agents, the technological advancements alone certainly cannot guarantee the user adoption of such interfaces. Hence, it seems appropriate to ask what is known about the users of such systems. For example, a question to be addressed is about the potential benefits of simulating interpersonal scenarios in human interactions with computers. The discussion of the empirical studies that have been conducted in an attempt to answer this along with other related questions, as well as, the problems involved in the empirical evaluations of embodied conversational agents in my domain of interest – mobile guide applications, is presented in Chapter 4.

In the next chapter the theoretical foundation of embodied guide agents is presented and discussed.

Chapter 3 Theories around ECAs

Now that I have explained what an embodied conversational agent is, and what it should be in the domain of mobile guides, it is time to raise the question as to why I should create one, and most importantly why I should build one for mobile interfaces. A question to be addressed is about the advantages of such agents in stationary/mobile interfaces. In this chapter, I carefully examine the theories behind ECAs.

The first part of the Chapter (§3.1), gives some evidence about why users would choose a metaphor of interpersonal communication for interacting with a computer and discusses it in relation to the mobile arena.

The second part of the Chapter (§3.2), outlines a number of theories on the possible effects of ECAs on the user's cognition, and gives a more detailed model of human cognition (§3.2.2), to provide a more effective theoretical framework for understanding these effects. Since this framework can explain human cognition as existing solely 'inside' a person's head, and does not account for the physical surroundings in which cognition takes place, an extension theory is proposed (see embodied tenet of the distributed cognition theory in §3.2.3). Based on these theoretical notions, I also define what cognitive accessibility, usability and user experience is about. The current work is then discussed in the context of the emerging field of augmented cognition.

Finally, the third part of this Chapter (§3.3), details a number of assumptions around ECAs along with the arguments that the critiques of ECAs have presented against their use in computer interfaces.

3.1 Effects on social responses

In this section, the question whether ECAs can elicit social responses from people is examined. To answer this question, the "Computers-are-Social-Actors" (CASA) experiments are presented that provide evidence that people respond to computers as whole, as well as to ECAs in a social manner. Then, the implications of the CASA-

paradigm on the way people interact with computers are discussed and an assumption is drawn on the applicability of CASA-paradigm in the mobile field.

3.1.1 The "Computers are Social Actors" paradigm

The main reason to build an ECA is to mimic the face-to-face social exchanges between humans. In using this visual dimension of interaction, rather than the more traditional keyboard and mouse, users can interact with the computer without having to learn complicated and unnatural computer commands, but rather in a natural and intuitive way (Cassel & Stone 1999). A question that many ECA researchers have pondered for some time is whether they can freely apply the metaphor of human-human social behaviour to human-computer interaction. Although this remains a largely unanswered question, some encouraging evidence can be found in the work of Reeves and Nass; "Computers- are-Social-Actors" (CASA) (Reeves & Nass 1996).

The researchers hypothesized that people treat media as if they were human. In a series of experiments, they tested this hypothesis by examining the extent to which humans respond to computers as if they were humans. Their findings show that people subconsciously respond to computers in a social manner, much as they do with other people. These social responses do not arise from conscious beliefs that computers are humans, neither are they the results of users' psychological or social dysfunctions, nor are they a result of a belief that users are interacting with programmers (Nass *et al.* 1994). Rather, social responses to computers are commonplace and easy to generate, even in circumstances where the users state that such responses are inappropriate. Therefore, the question arises what are the causes of the confusion of real-life and media, even though at any given time people are consciously aware of the fact that they are working with a computer. It might be assumed that a subconscious function is responsible, but the cause may be found in other factors as well. The answer of Reeves and Nass is short and clear; the human brain evolved in a world in which all perceived objects were real; humans are not adapted to distinguish between real-life and artificial media.

3.1.2 Can the CASA-paradigm be applied to ECAs?

As a continuation of the studies described in the previous section, Reeves and Nass (Lee & Nass 1988; Rickenberg & Reeves 2000) investigated whether the CASA-paradigm could be applied not only to computers as a whole, but also to Embodied Conversational Agents. Below I describe two of their experiments:

In the first experiment, the effects of social identification with respect to in-and out-group differences were investigated with ECAs. When meeting someone new, people categorize almost automatically that person as belonging to their "in-group" or "out-group", based on observable physical cues such as ethnicity and others. Computers do not have ethnicity for example, and for an ECA ethnicity is arbitrary. But based on the CASA model, I would predict the same responses to ethnically-matching ECAs as people tend to direct towards ethnically-matching humans. Indeed, in an experiment two groups of users consisting of males with Korean ethnicity had to report on choice-dilemma scenarios to two types of ECAs (one with a Caucasian and another with Korean appearance), which gave a scripted answer. Immediately after, participants were asked to indicate which of the agent's answer was similar to their own decision. Participants who worked with the Korean-like agent (i.e., the in-group agent) compared to those who worked with the Caucasian agent (i.e., the out-group agent), perceived it to have made a more similar decision, elicited more conformity to their opinion, being more attractive and trustworthy and presented more persuasive and better arguments.

The goal of the second experiment was to investigate whether users responded to the presence of ECAs in the same manner as they would respond to the presence of real people. Two groups of people were separated according to their "locus of control": people with an internal control (i.e., users who believe that the outcome of an event can only be determined by their own actions), and those with an external control (i.e., users who believe that the outcome of an event is determined by outside forces), were asked to perform a task. The first group was monitored by an ECA, while the second was not monitored by an ECA. Results supported the CASA-paradigm; the effects of monitoring and individual differences with regard to perception of control work as they do in real life. Users feel more anxious about their

work, and perform less well when an ECA watches, and this effect is stronger in users who think that other people control their success. Therefore, users perceive being monitored by an ECA as having the same effects, as those when being monitored by a real-human.

The findings of these studies provide strong evidence of the applicability of the CASA - paradigm to ECAs. Furthermore, more recent studies (Xiao 2006) have provided further evidence that reinforces this notion. Although, this is encouraging evidence towards the general adoption of ECAs; it tells us nothing about whether their use on computer interfaces can actually improve human-computer interaction. To the best of my knowledge, the current CASA-related studies have been made with stationary computer systems, where users are more or less isolated from the rest of the world. A mobile user, who is constantly interacting with other people and his/her surroundings, more likely may not experience the same social responses towards an ECA. For my purposes, however, I will not rule out the possibility that the CASA-findings may also apply to mobile ECAs.

3.2 Effects on Cognitive Functions

The use of ECAs in computer interfaces has been assumed to impact the ways in which people process information. In an effort to understand these effects a number of theories of human cognition are presented in detailed below. From those, the Simplex II model (Adams & Langdon 2003) is chosen as the theoretical foundation of this research work. The theory of distributed cognition (Hollan *et al.* (2000) is proposed as an extension to this model to match the requirements of mobile environments. In this section, the terms accessibility, usability and user experience are defined in the context of the current research work. Finally, a discussion is made on the possible utility of ECAs in the context of augmented cognition (Schmorrow & Kruse 2004).

3.2.1 Effects on Information Processing

Another part of the theory behind the ECAs is their effects on information processing. In the early nineties, a substantial amount of research was conducted on the impact of multimedia software on the users (Najjar 1996). There are two opposing

views with regard to information processing in multimedia environments. The first view, argues that multimedia environments cause information processing to be hampered; the user gets too much information, leading to cognitive overload. The other view suggests that multimodal input in semantically equivalent pieces of information contributes to deeper coding of information, most likely resulting in deeper learning.

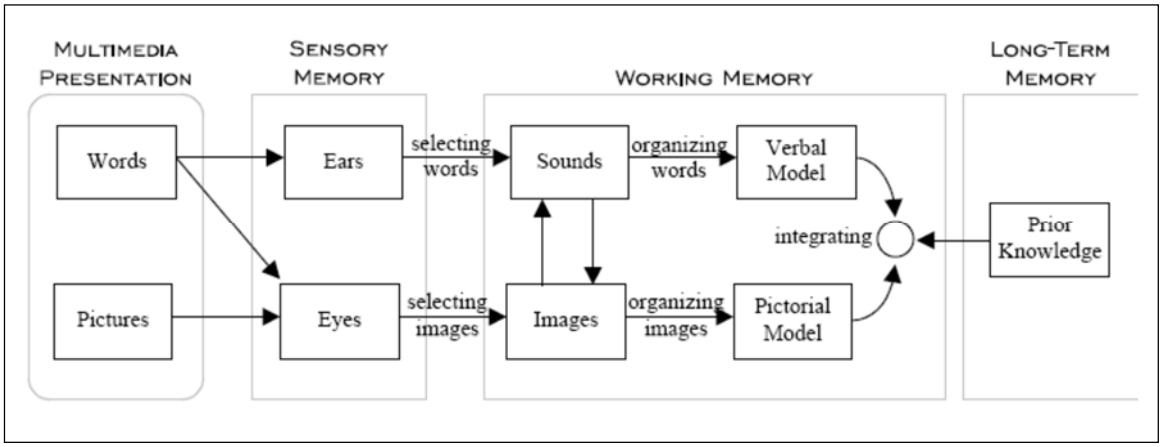


Figure 3.1: The Cognitive Theory of Multimedia Learning (source: Mayer & Moreno 2002)

Mayer and Moreno (2002) presented a cognitive theory of multimedia learning, which is based upon three primary assumptions:

1. Dual channel assumption: the idea that humans process visual and auditory representations/information through separate processing channels.

2. Limited capacity assumptions: the idea that each processing channel can actively process only a limited amount of representations/information at any given time.

3. Active processing: the idea that learning occurs when the learner engages in active cognitive processes such as selecting, organizing and integrating with existing knowledge.

According to this theory, people integrate both visual and auditory information into existing knowledge, but this process is most likely to occur when the corresponding pictorial and verbal information is present in the user's working memory at the same time (see figure 3.1).

With regard to ECAs, Moreno and Mayer state that people remember information better when it is encoded in a conversational, rather than a formal style. In an empirical evaluation with a pedagogical ECA (Moreno & Mayer 2000), the researchers showed that users work harder to understand information related personally to them, which results in deeper learning. This finding is perhaps to be expected as people require less cognitive effort to process personalized messages in their everyday conversations.

However, based on the above discussion it is not clear whether it is necessary to visually embody an agent, if only the conversational style is adequate enough to facilitate the integration of new information. To address this question, I need to consider the notion of "redundancy", and a number of research findings from cognitive psychology that clearly show that embodiment plays an important role in augmenting the facilitation of comprehension.

Redundancy can be defined as the quantity of presentation of identical information, provided simultaneously from different communication channels (for instance, the document in the below figure (figure 3.2) is presented in two different formats (one textual and one photographic) but both conveying the same information).

Copy

A. C. Barnes,
Merion, Pa.

New York , Sept. 20, 1930

Shall be at Savoy Plaza, New York City, September 28th until 30th. Would greatly appreciate permission to visit Barnes Foundation and meet you. Will you kindly let me know here at the Biltmore Hotel.

(Signed) Henri Matisse.

Figure 3.2: A sample document and its transcription (source: Corbis 1995)

Comprehension is directly affected by redundancy and previous mental models (or mental scaffolds) about how a system works, since it is more likely that the information being provided will be understood as redundancy increases. For example, if the information provided by one channel results in confusion and misunderstanding, then this channel can be supplemented by providing the same material through another channel at the same time (or approximately the same time) (Nemetz & Johnson 1998). Apparently, the strength of multimodal communication in ECAs, is exactly this; the redundancy of information transmitted through different communication modalities/channels. Additionally humans have perfected communication through those channels over thousands of years, ECAs, at least theoretically, should provide an intuitive interface between the user and computer that facilitates comprehension, without requiring any special user-processing capacity. This is in accordance with Lang (1995), who argues that the information provided by a talking head should require a fairly small amount of user processing, as humans are acquainted with processing verbal information spoken by a person.

Chawla *et al.* (1996) hypothesized that the conversational cues provided by a talking face, are sufficient to cause enhanced attention and thus improve recall of narratives in audio-visual environments. These cues might be important in increasing peoples' overall level of orientation and arousal[14] responses, and also could assist in processing verbal material by boosting speech comprehension[15]. Although, their study results show that the facial cues alone cannot enhance recall of narratives, these cues work in parallel with other conversational visual cues (such as static information like a person's clothes, face and physique, and dynamic information like movements of their hands, lips, head and eyes) in producing memory effects.

[14] It has been shown that the human face can attract stronger and longer visual orientation and tracking responses from infants, than other visual stimuli (Morton & Johnson 1991). A similar response may occur
in adults too, leading them to process narratives better when they see the speaker's face.(Chawla *et al.*1996)

[15] It has been demonstrated that the presence of visual articulatory cues (i.e., speaker's lip movements) can improve performance in speech shadowing (voice-writing) and serial recall tasks (Sumby & Pollack 1954)

3.2.2 A more holistic view of human cognition; The Simplex Two theory

Investigating the potential effects of ECA on the user's cognition requires a more analytical architecture of human cognition, than the model of multimedia learning can provide. I have chosen the Simplex Two (see figure 3.3) model of human cognition (Adams & Langdon 2003), for two reasons: First, it draws upon a considerable weight of research evidence (Adams 2006). Second, it captures a comprehensive understanding of human cognition, but it is still straightforward enough to guide system/ECA design, evaluation and application. The model postulates nine zones or modules of intelligent human behaviour, each of which can act partially independently of each other.

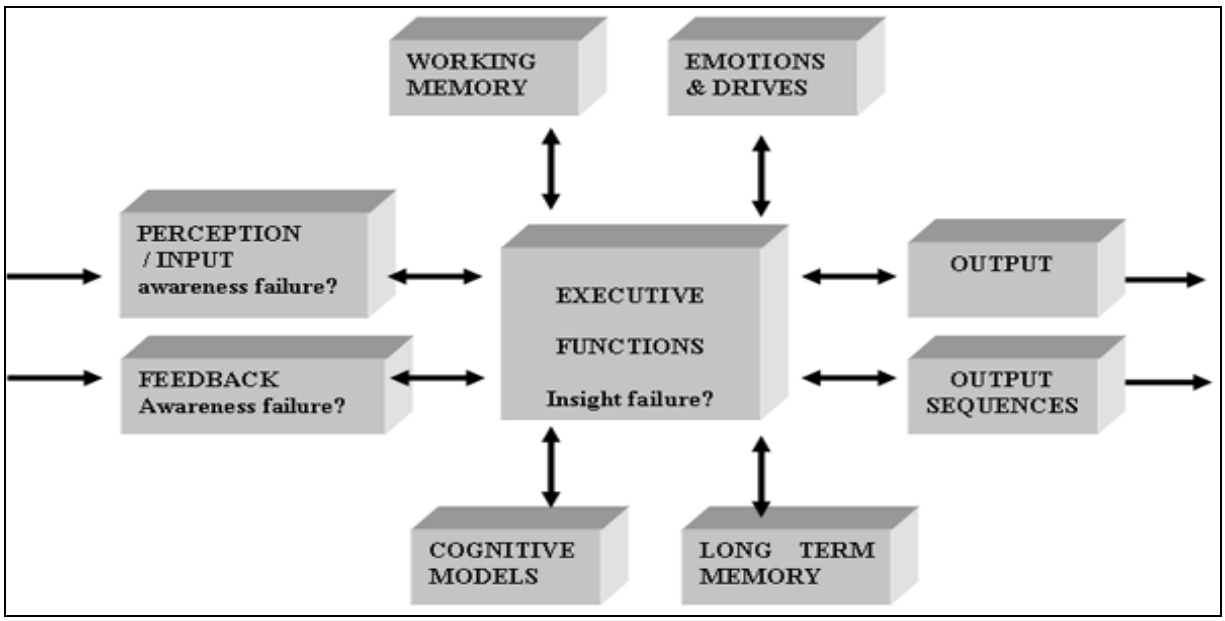


Figure 3.3: A Depiction of Simplex Two (source: Adams 2005b)

Each module has implications for the functionality and cognitive accessibility of an interactive system design, as well as for the user's psychology. In order to aid the understanding of the model, I present each of these modules/zones below, along with a discussion on the cognitive accessibility implications of each module for evaluating a design (Adams 2005a):

One: Executive Functions Module

This module reflects the general organization and implementation requirements of any cognitive task undertaken. It acts as a connecting node, transferring and transforming information between the other modules, according to the demands of the task and upon prior experience. A design is not accessible enough when it puts excessive demands to the users in terms of task complexity, learnability, inconsistency and self-organization.

Two: Perception/Input Module

This input store takes sensory and perceptual information from the environment into the system. It holds, evaluates, maintains and rehearses information. A design is not accessible enough when, given the sensory capabilities and skills of the people who are using it, its input modalities (visual, auditory or others) are not adequate enough. A system may be inaccessible, when its immediate memory requirements are too much for the user. An example of when this occurs when users are required to hold a complex visual or auditory display in mind while using a system. A system that is badly designed may not consider the perceptual skills of its intended users well enough, in order to avoid misunderstandings of the system's output to them.

Three: Feedback Management Module

This module deals specifically with the required feedback that a user needs to use when working with a system. A system is not accessible enough when the system provides inadequate feedback in terms of sensory (too brief or too loud), timing, relevance and memory requirements.

Four: Working Memory Module

This module operates as a working memory, providing, evaluating and transferring the required information, to support implementing a working task. Its storage capacity is limited and can degrade in efficiency over time or with disuse. A design is not accessible enough when it has too many demands from the user's working memory,

requires too much information to be held in the mind or the completion of complex operations.

Five: Emotions and Drives Module

This module generates the emotional responses to events and helps to relate cognitive and affective events to each other. A design that is not accessible enough, will receive inappropriate emotional responses. For example, the use of a system may prove to be a frustrating or annoying experience or not be serious enough for the task at hand.

Six: Long Term Memory Module

The warehouse of long-term associative memory holds the knowledge and expertise that needs to be retained on a long-term basis, including declarative knowledge (i.e., being able to state facts) and procedural knowledge (i.e., being able to carry out skilled actions). Capacity is very large, but relevant information may not be retrievable under all circumstances and conditions or it can even become permanently lost. On the other hand, the information can be reconfigured when the person carries out new tasks for long enough. A design is not accessible enough when it places too many demands in the user's long-term memory, in terms of knowledge or information acquisition, or provides inadequate support for long term learning when necessary, or does not provide adequate information to help retrieve information from memory.

Seven: Cognitive Model Construction Module

This module acts as a facility for building, retaining and applying mental models or maps to solve practical problems, such as navigation. Like the rest of the modules, this module stores models and processes them. A design is not accessible enough when it is not based upon a cohesive and understandable metaphor or a model, is not presented in an effective and accessible manner or makes it too difficult for the people who use it to acquire the structures involved and to model them.

Eight: Output Module

This module retains and generates responses, as a consequence of the work of the other modules and habitual responses, based upon triggers in the environment. A design is not accessible enough when the system requires the user to respond unreasonably, it provides insufficient respond support, or its requirements do not match with the capabilities and skills of its intended users.

Nine: Output Sequences Module

This module augments the other output module, by supporting the development of complex response based skills. For example, consider a part in a play or a sequence of moves for a dance or a sport. A design is not accessible enough when the system requires the user to respond unreasonably, it provides insufficient support for the development of user responses, or its requirements do not match with the capabilities and skills of its intended users.

For each component, I need to ask if the design of the mobile guide system is accessible enough. In this context accessibility refers to the degree to which a user can cognitively access, interpret and respond to the information (e.g., navigational information) conveyed by an ECA through speech and body. The potential cognitive barriers that an ECA may pose to the user include:

- Verbal and non-verbal channels are out of synchronization: Body movements are out of synchronization with the ECA speech, disrupting the information conveyed in speech, and hence, causing confusion to the user.
- Non-cultural oriented hand gestures: Certain gestures have different meaning to different countries. Their wrong use can disrupt the conveyed information and cause confusion to the user.
- Speech quality: Certain voice attributes (e.g., too loud, too soft or gaps while speaking) might make it hard for the user to actually understand what the ECA is saying.

- Various aesthetic attributes: Certain aesthetic attributes (e.g., the ECA's dressing, hair, face, etc.,) may create feelings of discomfort to the user that may limit his/her capacity to process and respond to the conveyed information.

In parallel with my investigations on the potential impact of an ECA on the user's ability to cognitively access the information contents of a mobile guide system, I could also account for any potential effects on the *usability* of the system. In this context of use, usability refers to the extent to which a user can interact with an ECA to navigate, and extract information of interest from a particular environment with *effectiveness* (e.g., whether the user did a better job in the tasks with the ECA, than without it), *efficiency* (e.g., how quickly was the user able to complete the tasks with the ECA) and *satisfaction* (ISO 9241-11 (1998))

From the above discussion, it is clear that cognitive accessibility is a pre-condition for usability, but not vice-versa. If the ECA hampers the user's ability to cognitively access the information contents of a mobile guide system, it is unlikely that s/he will be able to actually use the system to achieve the assigned tasks. From this point of view, one could also argue that usability is a subset of cognitive accessibility, as even routine tasks require some degree of cognitive processing to complete successfully. Despite this strong relationship though, each term refers to different aspects of the interaction process and hence, should be kept and evaluated separately.

3.2.3 Theory of Distributed Cognition

A theoretical framework for research in mobile ECAs would have been incomplete without considering cognition beyond the individual (as illustrated by Simplex II) and to the environment as a whole. In such settings, there is the need to extend what is considered cognitive to encompass interactions between people and resources and materials in the environment. From the existing HCI theories, distributed cognition (DCog) (Hollan *et al.* (2000) is a suitable theory to understand these interactions in the current context. As opposed to other theories, like for example Activity theory (Kaptelinin 1995), DCog has three basic principles (i.e., socially distributed cognition, embodied cognition, and culture & cognition) that are easy to understand, and can be

adapted (with minimal effort) to ECAs. Each of these principles is elaborated below, along with assumptions/arguments about the possible effects/implications of ECAs:

The basic idea behind the concept of socially distributed cognition is that cognitive processes are distributed across the members of a group, according to the group's social organization. Its argument is that cognitive processes involve the transmission and transformation of information across several pathways, and when the patterns of these pathways are stable, they reflect some form of underlying cognitive process. Since social organization – plus the structure added by the context of activity – is by itself a deterministic factor of the information that flows through a group, social organization may be considered to be a form of cognitive process. The broader conception of socially distributed cognition includes, in addition, phenomena that occur in social interactions as well as interactions between people and structures in their environments.

With regard to ECAs, some arguments can be found in the literature that suggest they could have a positive impact on a group's cognitive processes, for example, by facilitating the communication between the group members (Kruppa *et al.* 2005). In particular, Kruppa *et al.* (2005) argue, that the social supportive aspects of ECAs could encourage collaborative learning within the members of small groups, when visiting a museum. In order to evaluate this potential, they propose a comparative user study from which they expect that the groups accompanied by a version of their system with an ECA will have a lower perception of task difficulty and higher knowledge gain than those without an ECA. I share this belief, but I do not plan to address it in the current research. As my work focuses on the cognition of a single person, this social aspect of distributed cognition will remain an open question, hoping that other researchers will find it interesting for future empirical work.

The second principle of the distributed cognition approach is that cognition is embodied. It is not incidental that we have bodies and we use them for causally linking with our immediate environments. This relation is an essential fact of cognition that evolution has designed us to exploit. In other words, this approach postulates that the human mind is not a passive representational engine whose only function is to create internal models of the external world. There is a closer and far

more complex relation between internal and external processes, which involves coordination at many different time scales, between internal resources such as memory and attention, and external resources such as the objects and artefacts, constantly surrounding us.

A crucial moment within this coordination, that can decide its success, is whenever the user needs to switch his/her attention focus from the virtual world of the mobile device, to the actual objects/artefacts of the physical world. An ECA could be used to effectively guide the user's attention focus towards these physical objects by helping the user in understanding the underlying structure of the physical space. It may, for example, help the user to locate a particular landmark by directing his/her gaze (through the use of proper gestures) towards its current location in the virtual world. This should allow the user to better locate the object in the physical environment. In the very same way, the ECA may help the user to identify parts of the current structure/object s/he is looking at, and for which the system is providing information for. I have investigated this aspect of ECAs in a series of exploratory studies under simulated mobile conditions (see Chapter 7 and Chapter 8).

The third principle of the theory of distributed cognition holds that cognition is not isolated or separable from culture, because people live in complex cultural environments. In this perspective, culture not only emerges out of the activities of individuals (as cognitive anthropologists traditionally believe), but it also shapes cognitive processes of systems – those that are distributed over people, artefacts and environments. The basic notion behind this linkage is that the environment within which people are embedded poses significant challenges for learning, problem-solving, and reasoning. Culture is a continual process for accumulating solutions to frequently encountered problems. This residue of past experiences is a reservoir of resources, from which humans could begin looking for solutions, informed by the mistakes or successes of others. For this reason, culture provides us with the intellectual tools to acquire this knowledge, and accomplish things that we would not do without them.

My view of ECAs in this context is that humans seek the same cultural traits (verbal and non-verbal behaviours, appearance, etc.) in their communication with the agent as

they would in human-human communication. Given the fact that humans prefer other humans who are closest to their own demographic variables (and given that that culture could affect cognition) then care should be taken in designing the ECA in such a way to match the cultural characteristics of the intended group of users. For example, some evidence can be found in the literature that users, when given a choice to select a character from a pool of characters with different ethnicities to work within a cooperative task, they would select a character that matches their own ethnicity (Cowell & Stanney 2002). Since this is a fact in all human cultural environments, I see no reason why this finding could not be applied to characters in mobile environments as well. Although this assumption is interesting, I do not intend to include its investigation in my current research.

Having included the embodied principle of distributed cognition in my theoretical framework, the terms of cognitive accessibility and usability (which I introduced in the previous section) are no longer sufficient to cover users and their interactions in whole environments. A more holistic, all-encompassing notion is needed to describe effectively those settings. The user experience, that is the overall experience a user gets when s/he interacts with an ECA in particular conditions, fully satisfies this description. More specifically, user experience refers to the cognitive accessibility and usability of an ECA interface, and includes a number of factors that influence the experience the interaction evokes. I have identified and seek to investigate the following key factors: 1) different features of the user, 2) numerous aspects of how the ECA looks, sounds, and behaves, 3) context of use, and 4) different task features. These factors, along with a number of variables that can be empirically investigated, are discussed in detail in the next chapter (see Chapter 4).

3.2.4 The Notion of Augmented Cognition

Augmented cognition is an emerging field of science that seeks to extend the information processing capacity of individuals working with 21st century computing technologies via computational methods and tools explicitly designed for accommodating bottlenecks, limitations, and biases inherent in human cognition (e.g., limitations in attention, memory, learning, comprehension, visualization abilities, etc.) (Schmorrow & Kruse 2004). This notion becomes of particular importance, when it

comes to mobile environments, where the cognitive limitations, of users become more severe when compared with a stationary set up. The use of ECAs in mobile interfaces holds a high potential of compensating for these limitations by rendering the interactions multimodal. With an ECA, a mobile system can rely on all sensory and motor processing channels in addition to natural language, for receiving and conveying information. In this way, the use of an anthropomorphic interface might augment the user's ability to interact with such systems easily and naturally in order to achieve their goals. In the current mobile guide scenario, this means that the ECA should potentially augment the participant's ability to navigate the physical environment, as well as to retain effectively and efficiently the information provided by the system. However the ECAs should not be evaluated solely for their potential effects on the user's cognition. As elaborated in the next section, ECAs can affect users in other ways rather than just purely cognitively.

3.3 Other Assumptions about ECAs

Being able to communicate with a computer in a social manner is said to lead to a more engaging and motivating human-computer interaction. This should in turn support cognitive functions such as problem-solving, understanding and learning (Van Mulken *et al.* (1998)). According to Lester *et al.* (1997), ECAs can more actively engage students in learning, which may well stimulate reflection and self-explanation[16], thus resulting in enhanced learning performance. In addition, their mere presence in an interactive learning environment can motivate students to perceive their learning experiences more positively.

One other aspect where ECAs are believed to have a positive impact is the ease with which users learn the system functionality (i.e., learnability/smoothness of interaction). Rendering the computer interface more human-like and socially realistic, should make interacting with it smoother, as humans already know how to interact socially. Interacting socially should not only be easier, but also more enjoyable than ordinary computer interactions. In mobile guide tourist systems every effort should be

[16] Self-explanation is the process where a student explains after reading a line of text, his/her understanding of the text, and a piece of generated knowledge that goes beyond what the sentence explicitly said. (LearnLab Wiki, 2013)

made to make the user comfortable, and his /her experience with the system as enjoyable as possible. If so, the presence of ECAs in those interfaces should therefore improve their effectiveness and enable the realization of interactive and personalized tour experiences.

The use of ECAs in computer interfaces could also influence the ways in which users perceive the believability of the system. Users must find the information provided by the system to be convincing. As people enjoy interacting with characters all their lives in films, books and on television, an ECA seems as an ideal candidate to achieve believability in a system. However, as Dehn and van Mulken (2000) found, a person is more likely to attribute a greater degree of believability to a system if it is perceived as both intelligent and competent. As I have already discussed (see §2.2.2 of Chapter 2), believability is an issue of some controversy between the two major lines of research in the ECA field: Cassel and her group, who assert that implementing more natural conversational functions into ECAs will result in higher believability (Cassel & Stone 1999), and researchers like Bates (1994), and Hayes-Roth and Doyle (1998), who believe that there is more to learn from art than from biology.

Very much correlated with believability is the aspect of trust. The formation of trust with users is crucial in all systems that provide information and when users are expected to interact with the system for a long period of time. Cassel's approach in building trust in ECAs is again realistic; human strategies for establishing and maintaining social relationships where trust is important can be successfully used to realize the same results for ECAs. She argues that interaction rituals among humans, such as greetings, small talk and conventional leave-takings, along with their manifestations in speech and in embodied conversational behaviours, can lead the users of technology to judge the technology as more reliable, competent and knowledgeable and hence, to trust the technology more (Cassel & Bickmore 2000). Also, those interaction rituals ECAs can demonstrate the influence of benevolence by relating past experiences of benevolent behaviour or referring to third-party affiliations.

The realistic approach does not seem to work for building relationships over a period of time. Bickmore and Picard (2004), argue that their experience with the REA[17] system showed that current state-of-the-art in speech recognition and natural language understanding does not come close to supporting "the social dialogue (and conversational speech register) required for long-term relationship building". Even in a Wizard-of-Oz scenario[18], subjects found that the fixed repertoire of REA's responses left them feeling that REA was not really listening to them. For this reason, their FitTrack system uses scripted utterances in a menu like format, where users select from available options. Feedback from the system's evaluation indicated that most subjects found the dialogue, for both social and health-related issues, to be natural and fluid.

A somewhat different approach to portraying trust in ECAs is the one of Cowell and Stanney (2003) that uses nonverbal behaviours. According to them, specific elements of human-to-human nonverbal behaviour can help portray successfully a credible façade for an ECA. They suggest a number of recommendations for the design of credible nonverbal behaviour in ECAs, based on a taxonomy which organizes nonverbal behaviours into functional categories (i.e., facial expression, eye contact, paralanguage, gesture and posture) and the manner in which they can be embodied, based on literature suggestions from several fields (i.e., sociology, psychology, social psychology and political science). A number of prototype characters were created based on these recommendations and were empirically evaluated. The results showed that such behaviours may indeed lead to the creation of a more trusting environment and more positive experiences for the users.

Regardless of the potential advantages of ECAs and the continuous effort to improve them, there are several researchers who view them as impractical and inappropriate. In the next section, I review the main arguments presented by people opposing the use of ECAs in machine interfaces.

[17] REA (Real Estate Agent) is a lifesize virtual human capable of multimodal input understanding and output generation in the domain of real-estate (Cassel *et al.* 1999). She is one of the most well-known examples of ECAs.
[18] Natural Language Understanding in Wizard-of-Oz setups is performed by a human confederate

3.4 Opponents of ECAs

Because ECAs introduce a whole new paradigm of interacting with computers, there are a lot of people from the field of HCI-research who criticize them. ECAs violate a lot of HCI principles and hence, many researchers believe that they hamper human-computer interaction. In particular:

One argument focuses on the possibility that humanized interfaces may induce false mental models of the system (Shneiderman & Maes 1997, Norman 1994). For instance, the human-like behaviour of an agent, may lead the user to generalise the system's behaviour, and believe that it imitates humans in other cognitive aspects as well. Because of this, the user may be led to expect capabilities of the system that it does not actually possess.

A second criticism expressed by Shneiderman (one the fiercest critics of ECAs today) (Shneiderman & Maes 1997), is that anthropomorphic representations may mislead both designers and users, increase anxiety about computer usage, interfere with predictability, reduce user control, undermine users' responsibility and destroy the users' sense of accomplishment. However as Maes (Shneiderman & Maes 1997) argues, ECAs should not act as a substitute for a user interface, but rather complement it. This would mean that, in a given application, the agent should allow the user to interact with the application's interface and intervene only when the user requests assistance.

Wilson (1997) points out that, unlike conventional graphical user interfaces with buttons and menus where users can easily construct a mental model of possible actions (every button maps to only one action), interaction with ECAs does not permit the construction of a user mental model so easily. When interacting with ECAs, there are no buttons to indicate what the user can or cannot do with the agent; users will have to find out the agent's abilities or limitations by themselves. Therefore, Wilson stresses that a simple cartoon representation in visualization is preferred, rather than more realistic depictions. As an ECA becomes more realistic the less it is allowed to exaggerate or break the rules of the natural world in the user's mental model.

A further problem with anthropomorphization arises when the animations used to render the agent do not map onto the system's behaviour. For example, there are systems in which inactivity is visualised by an agent doing the so-called idle-time movements, like tapping its foot, etc. This however, may be falsely interpreted by the user for system activity, something which might lead to less efficient interactions with the system (Wilson 1997).

Anthropomorphising an agent can also lead to cognitive overload. Opponents of ECAs, argue that the presence of an additional (eye-catching) object on the interface might put additional demands on the user's cognitive resources by constituting another potential source of distraction (Walker *et al.* 1994). This assumption is partly supported by empirical evidence (Wright *et al.* 1999), which suggests that animated graphics in electronic documents may impair or distort user performance in text retention.

It might be not necessary to anthropomorphize computer interfaces. According to the CASA-paradigm, users will exhibit social behaviours towards the computer even if it is not visually embodied as a human being. Minimal social cues, such as a human sounding voice or simple language output, can induce the user to elicit a wide range of social behaviours. Therefore the use of more complex features might not be necessary.

Finally, there are the philosophical and moral concerns associated with the use of ECAs on computer interfaces. At the philosophical level we are still redefining what it means to be human. As Shneiderman (1997) says, *"people are not machines and machines are not people"*. The inability to distinguish humans from the machines, the real from the artificial, or to be able to terminate the machine by simply "pulling the plug", plagues the world of science fiction. According to Haraway (1991), *"late twentieth-century machines have made thoroughly ambiguous the difference between natural and artificial, mind and body, self-developing and externally designed, and many other distinctions that used to apply to organisms and machines. Our machines are disturbingly lively, and we ourselves frighteningly inert"*. At a moral level there is the question of whether it is ethical to use artificial means of human-to-human communication to trick users in a simulated relationship. As Massaro *et al.* (2000)

argue ECAs could help minimize the negative effect of machines on a user's depression by simulating human communication and building of relationships. However, if the agent suddenly stops functioning for a period of time (or even indefinitely) a depressed/lonely user, who has developed a close relationship with the agent, could most likely react badly. This creates the ethical dilemma whether it is appropriate to misuse the user's feelings in a more forceful way than existing methods already do (e.g., in video-games and television).

Summarizing, arguments in favour of ECAs are:

- ECAs can improve certain cognitive functions through enhanced motivation.
- ECAs can positively affect learnability.
- ECAs can positively affect the believability of a system.
- ECAs can enhance trust-building with a user.

The main arguments against ECAs are:

- ECAs can induce false mental models of a system.
- ECAs can reduce the sense of user control.
- ECAs might lead to cognitive overload and hamper user performance by distracting the user from the task.
- ECAs are not perhaps necessary, users respond socially to computers with minimal social cues.
- It is unethical to trick users into simulated relationships with ECAs.

3.5 Conclusion

A number of theories behind ECAs were presented in this chapter, along with a number of assumptions as to why their use on computer interfaces would be effective. From the theories discussed, the Simplex Two theory and the embodied principle of distributed cognition were chosen as the theoretical foundation for the current research. The Simplex Two theory is a simple architecture of human cognition, that can provide a powerful theoretical framework upon which investigations of the

potential effects of ECAs on the user's cognition can be built, but it only supports cognition for the individual human mind. The embodied principle of distributed cognition fills this gap, by extending the reach of cognition to encompass interactions in which the user pursues his/her goals in collaboration with elements of the material world surrounding him/her. Based on this amalgamation of theories, I also explain some controversial terms in HCI, i.e., what cognitive accessibility, usability and user experience mean, at least in the context of this research work.

A discussion of the empirical studies that have been conducted to test mostly the assumptions discussed in this chapter is presented in Chapter 4.

Chapter 4 Related Experimental Work

This chapter reviews relevant empirical studies that were conducted to evaluate the potential impact of ECAs on the users of mobile and mobile tour guide applications, along with some selected studies from the wider literature on ECAs for stationary environments.

The results of the evaluations are discussed and critiqued, and an attempt is made to answer the question whether the use of ECAs on the interface (mobile and stationary) provides any advantage over not using them. Following the problems that the literature leaves open for mobile interfaces and mobile tour guide interfaces are elaborated, a review is reported about the limitations of the current research in the area. Finally, a number of appropriate questions and hypotheses are generated to be studied in this research work.

4.1 Studies on Cognitive Functions

As discussed in the previous chapter (see §3.2 of Chapter 3), by being able to communicate with a computer as one does normally in social interaction with other people, would support (at least theoretically) cognitive functions such as problem-solving understanding and learning. Although there is a growing pool of empirical data relevant to these effects, very little of this work actually refers to mobile environments. Below, I discuss a number of studies from the realm of mobile devices, and a few selected studies in ECAs for stationary devices.

A study in which the impact of a mobile affective guide was evaluated on the users' recall performance was performed by Lim and Aylett (2007). Three different types of mobile guide were evaluated in an interactive tour of the "Los Alamos" site of the Manhattan project. All stories generated by the system, related to the "Making of the atomic bomb". The physical tour, however, took place at the Heriot-Watt Edinburgh campus buildings, where buildings from the "Los Alamos" site were mapped onto University buildings. The agents differed in terms of emotions and attitude, portrayed by a simplistic 2D cartoon-like head, and by the inclusion of the agent's perspective

and experiences in the narration. The fully affective guide could exhibit both emotions and attitude while the other two displayed no emotions nor attitude, or emotions but no attitude respectively. For example, the emotional guides could dynamically update the story to include their own feelings and perspective about historical facts (e.g., *"it seemed brutal to be talking about burning homes"*). The participants were requested to listen to at least three stories at each location, under a thematic area of their choice (e.g., Science or Military). After the completion of each story, participants had to rate the degree of interest of the stories, as well as how much they agreed with the guide's argument. In the first group (i.e., the fully affective guide), the input given influenced the processing conducted by the guide, while in the other two it merely gave the impression that it did. Upon completion of the tour, participants had to answer two sets of questionnaires, one to indicate their subjective experience of the system, and another to test their recall levels of the information they listened during the tour. In terms of recall performance, the researchers found no significant differences in the users' recall levels of the presented information between the three guides. They attributed this to various confounding variables, such as the speed of the guide's voice, non-native English speaking users, etc.

I find this non-effect result to be rather expected. Although the conclusion that a guide with attitude and intelligence makes the interaction more interesting may be a valid conclusion, it is how these behavioural attributes are portrayed through non-verbal means that can translate subjective views into enhanced retention performance. For example, studies have shown that a more realistic depiction of a virtual human (MacDorman *et al.* 2010) can create greater participant involvement in a virtual experience. In addition, the absence of body language (e.g., beat gestures) deprived the guide of a valuable communication channel (see the notion "redundancy" in §3.2.1 of Chapter 3) that would augment the presented information and, in turn, lead to greater user retention performance. Last and perhaps most importantly, the pretended Los Alamos site at the University campus made it impossible for the users to connect the content of the stories with their surrounding environment, thus causing unnecessary cognitive overload (see the second principle of distributed cognition in §3.2.3 of Chapter 3). If the user's cognitive resources were devoted into blocking the external stimuli in order to focus his/her attention to the stories, it is not surprising that the attitude and emotions of the agent had no impact on his/her overall retention

performance. If the study had been conducted on the actual site, where users would have the ability to physically visit the various sites, the results may have been different.

Kruppa *et al.* (2005) suggested a study that would investigate the potential effects of ECAs on collaborative learning in small groups of visitors in museums. They suggest asking a group of 100 school-aged children to use the mobile museum guide, half with an ECA to browse the museum, and the other half without the ECA. After an initial screening phase the system automatically partitions each of the groups into experts and novices. Before asking participants to complete some collaborative tasks, the system further organises the participants into groups of three or four, based on their expertise in certain areas. After completing the tasks, each child fills-in a questionnaire on aspects of their learning experiences. The researchers expect that the groups with the virtual character will have a lower perception of task difficulty and higher knowledge gain than those without the virtual character. However, the outcome evaluations have yet to be reported, and thus this conjecture remains speculative.

In the domain of health care intervention, Bickmore (2007) investigated the impact of four different handheld ECA interruption strategies on the long-term health behaviour adherence of users. Four ECAs (each with a range of nonverbal conversational behaviour and appearance) used one of four interruption strategies to advise a user performing an office task about the importance of taking frequent breaks from typing in order to avoid repetitive stress injury. The four strategies evaluated were: NEGOTIATED; the ECA provided users with the ability to delay the start of an interruption via a snooze button; FOREWARN; the ECA gave users a warning that the interruption was about to occur; SOCIAL; the ECA apologizes for interrupting the user, and provides empathic feedback based on the user's emotional state at the time of interruption; and BASELINE; a simple audio alarm was used as a control condition to compare against the other three conditions. Results suggest that the users rested the longest overall in the SOCIAL condition, but the difference between the conditions was not statistically significant (possibly because of the small size of the group). Thus, the researchers concluded that the use of social behaviours, such as empathy, in recommendations for health behaviour change leads to better user compliance as

opposed to non-social interruption strategies. I suspect that this finding supports the use of ECAs in mobile tasks, with the goal of improved task performances as opposed to not having an ECA in the interface. While this result can be generalized only with caution (as it might be application or ECA specific), the use of empathy and other forms of social interaction in dialogues with ECAs, should at least increase the likelihood of improved task outcomes in other domains as well.

In the same domain, Johnson *et al*, (2004) developed DESIA, a psychological intervention ECA, to help teach problem-solving skills to caregivers of paediatric cancer patients. A formal evaluation was scheduled with, among other goals, the investigation of the impact of DESIA on the users' problem solving ability and general affectivity over time. The outcome of this evaluation, however, has yet to be reported.

Moving to ECAs in stationary environments, as was rather expected, there is a richer literature than for mobile ones. Furthermore, a number of these studies have revisited the results of older ones (Miksatko *et al*. 2010) and others have examined previously unexplored effects (e.g., Eichner *et al*. 2007).

Eichner *et al.* (2007) investigated the use of eye gaze as input to an agent-based virtual sales scenario. The system adapts the presentation to match the user's interests, and reacts appropriately if the user is inattentive. In an exploratory study, two versions of the system were compared: a fully interactive version which analysed the user gaze behaviour in real-time and provided appropriate reactions to interests/disinterest, and a pseudo-interactive version which was based on randomly assigned interruptions. They found that in the interactive version, the agents guided the user's attention more successfully to the content of the presentation (i.e., product images and slides) than in the pseudo-interactive version. However, this form of guidance made users more uncertain on which of the products to actually select (as this was indicated automatically by the system). Thus, it was concluded that user-aware agents can be used for successful guidance to the desired interface objects, but more work is needed to determine the optimal level of agent attentiveness. I have examined the impact of a basic user-aware agent, on the retention of information presented by an ECA in simulated mobile conditions (see §7.1 of Chapter 7), but the

impact of a user-aware ECA, similar to the one discussed above, remains to be investigated in both immobile and mobile environments.

Using a much less sophisticated system, Beun *et al.* (2003) investigated the effect of ECAs on the user's retention of information. Users experienced two small stories from three types of ECAs – a realistic female head, a cartoon purple gorilla, and an absent animated agent, represented by a word balloon – and were asked to write down everything they could remember from the narrations. The researchers found that the participants could remember significantly more in the "realistic" condition, than in the "absent" condition, but there was no statistically significant difference between the "realistic" and "cartoon" conditions. Hence, they concluded that the presence of an ECA had a positive effect on memory performance, but this effect does not depend on the degree of anthropomorphization of the ECA. However, this conclusion is not definitive. As the stories were the same across the "realistic" and "cartoon" conditions, the participants may have simply not noticed any differences between the characters. If the discourse was longer and was varied between the agents (e.g., serious vs. fun), the effect would have been most likely different. For example, it is a plausible hypothesis that for serious discourse participants would have remembered more in the "realistic" condition than in the "cartoon" condition. Then, a longer discourse would have demanded more attention in both conditions than the shorter discourses used in the experiment. The extra attention would have probably resulted in a non-significant difference between the "realistic" and "absent" conditions. This "Persona Zero-effect" is best demonstrated in the following study.

In an older study by van Mulken *et al.* (1998), designed to investigate the impact of a presence of an ECA on comprehension and recall, participants were presented with a series of technical and non-technical materials presented either by an ECA or an animated arrow symbol. The technical material consisted of information about different pulley systems, while the non-technical was an introduction about ten fictitious employees of a research centre. The results showed no significant difference between the ECA and graphical arrow conditions, neither for technical nor for non-technical information. Hence, the study concluded that adding an ECA has no detrimental effect on comprehension and recall, but, also, does not improve it, and that the type of information does not play any role in this.

The above study was recently revisited by Miksatko *et al.* (2010). One of the major differences was the repeated interactions. Specifically, the researchers evaluated the impact of an ECA on motivation and learning performance in a repeated task over a period of time. In the study, each group of participants experienced a vocabulary trainer application, either with an ECA (*with-agent* version) or without an ECA (*no-agent* version). In the no-agent version, the user interface consists of two windows displaying the English and German expressions and a row of buttons for showing and rating the answer. In the with-agent version, a female ECA was added in the middle of the screen featuring some idle movements to make her look alive and with a minimum amount of gestures. The researchers reiterated the findings of van Mulken *et al.* (1998) about the "Persona Zero-effect", i.e., that they found neither positive nor negative effects on motivation and learning performance. Therefore, the researchers concluded that adding an ECA on an interface does not benefit performance but also does not distract.

The no-effect results produced by the two studies above are, in fact, encouraging. If the mere presence of an ECA (with minimal or without nonverbal communicative behaviours) in a learning environment has no detrimental effects on performance, then it could be assumed that endowing ECAs with a full repertoire of proper nonverbal behaviours might, in fact, improve performance. Comprehension can be directly affected by redundancy (see §3.2.1 of Chapter 3) and, hence, the use of an additional, redundant channel of communication, such as gestures or facial expressions, could result in more learning. In particular, the use of gestures could reduce message ambiguity by focusing learner attention, and facial expression can reflect and emphasize the agent message, emotions, personality and other behaviour variables (Baylor *et al.* 2008). The study discussed below attempted to examine the validity of this hypothesis.

In a large scale user study, Baylor *et al.* (2008) investigated the effects of an ECA's non-verbal communication on attitudinal and procedural learning. They conducted an experiment during regular sections of an introductory computer literacy course in two separate stages, a procedural stage and an attitudinal stage. In both stages, each participant experienced one of the two types of course knowledge (procedural, attitudinal), with either an ECA's deictic gestures (presence or absence) or face

expressions (presence or absence). They found that facial expressions were perceived as more desirable for attitudinal learning, but they had a detrimental effect on procedural learning. On the other hand, deictic gestures were perceived more positively and lead to more learning (measured by recall), while learning procedural knowledge as opposed to acquiring attitudinal knowledge. The study concluded that designers should consider the type of knowledge they want to represent and transmit, and then decide the type of nonverbal animation that effectively enhances and augments the nature of the message.

4.2 Experiments on Other Assumptions about ECAs

The use of ECAs in computer interfaces does not affect users in a purely cognitive way, but it can also influence their attitudes towards the system and the task it supports. In addition, it is has also been speculated that an ECA can affect the user's style of interaction with the system. As before, while there is ample empirical evidence for stationary ECAs, the relevant studies in mobile environments are limited. Below for each category of effects, I first discuss and critique a number of studies for mobile and their findings, and then focus on a few relevant studies from the context of stationary computer applications.

4.2.1 The User's Subjective Experience

Lim and Aylet (2007) found that users who interacted with an emotional version of their storytelling tour guide ECA had a better overall tour experience and found the stories more interesting than those who interacted with a merely random emotions guide. In addition, the emotional guide was perceived as more believable in its discourse and its personality was described with more positive adjectives (e.g., interesting, helpful, funny, etc.), than the random emotions guide. In a similar fashion, participants in the Bickmore study (2007) rated the advice of the social version of a health advisor ECA as significantly more effective at getting them to rest when they are working at a computer, than the other three versions of the system: NEGOTIATED, FOREWARN and BASELINE (see §4.1 for more details). Finally, they perceived the social version as significantly more polite and more desirable for continuous use than the other three conditions.

This pattern of positive effects on the user's subjective impressions seems to be repeated in ECAs for immobile applications. In a study by Babu *et al.* (2007), an immersive virtual human training application (condition VR) in the domain of social conversation protocols was compared against instruction based on a written study guide with illustrations (condition L). In a post experimental evaluation, participants indicated that they generally enjoyed interacting with the virtual humans, found the multimodal interaction to be intuitive, and the instructions clear. In addition, they also viewed the system as useful for training verbal and nonverbal skills when required, and indicated that they would use the system frequently if it were made available to them. On the other hand, participants in the L condition found the written study not as useful as learning from example, practice and feedback. Although the ECAs used in this study would react to nonverbal input and generate appropriate nonverbal output, the mere use of nonverbal behaviours in ECAs can lead to positive user experiences as is demonstrated by the next study.

Kramer *et al.* (2007) investigated the effects of different nonverbal behaviours of an ECA on the users' experiences. Participants interacted with different versions of an agent whose nonverbal communication was manipulated with regard to eyebrow movements and self-touching gestures (presence or absence). Results suggest that self-touching gestures have positive effects on the experiences of the user as opposed to their absence (no self-touching). On the other hand, eyebrow raising evoked less positive experiences in contrast to no eyebrow raising. However, an interaction between the self-touching and eyebrow behaviours showed that participants perceived the agent more positively when it displayed both versus neither of the behaviours. Hence, the study concluded that the effect of specific cues may be dependent on the presence or absence of others and that more detailed research is needed to address these interactions. Although such research would be beneficial, it is also evident from this study that when different nonverbal channels are combined (facial and gestures), they can create positive user experiences regardless of the potential impact (positive or negative) of each channel to the user.

One of the most important aspects of the user's experience with any interface is trust. Gaining the user's trust is a fundamental factor towards a successful interaction. A number of studies have asked the question whether the social nature of ECAs can

assist the achieving of user trust in computer applications. Below I review some relevant studies from the relevant literature.

A study by Bickmore & Mauer (2006) on the perceived credibility of information delivered by four versions of a handheld ECA found no significant differences between the conditions. The four versions evaluated were: (FULL) a full version of the ECA (animation text and sound); (ANIM) an animated ECA without nonverbal behaviour; (IMAGE) an ECA showing a static image of the character; and (TEXT) the interface without any character. Although, the researchers do not provide an explanation for this lack of effect on the users' perceived credibility, it may be attributable to the relatively short time (i.e., five minutes) participants interacted with each version of the system. As in human social interaction, it typically takes more time to put trust in the information provided by an individual.

McBreen *et al.* (2001), found that in the context of three interactive electronic retail desktop applications - a cinema box office, a travel agency and a bank – participants rated the agents' trustworthiness significantly differently across the three applications. The qualitative results showed that participants were less likely to trust the agents to correctly complete the tasks in the travel and banking applications (with banking being the less trusted one) and more in the cinema application. Participants justified their preference by arguing that interactions in the banking and travel applications are more critical, with users becoming more anxious if something goes wrong. However, the researchers argue that responses to ECAs in more serious applications may be improved if the agent's trustworthiness can be established firmly. The study discussed below investigated the problem of portraying trust in ECAs in more depth.

The issue of trustworthiness/credibility was investigated in more depth in a study by Cowell and Stanney (2003). The researchers suggested the use of nonverbal behaviours as a method of portraying a credible façade for ECAs. Based upon recommendations from the literature for credible nonverbal behaviour, the researchers created and empirically evaluated a number of prototype desktop computer agents. The results obtained indicated that by including trustworthy nonverbal behaviours, the credibility of ECAs is enhanced, while the addition of trustworthy bodily nonverbal behaviours did not provide much credibility, over that created by trustworthy facial

nonverbal behaviours. Moreover, and perhaps most importantly, an ECA that expressed non-trusting nonverbal behaviours was thought to be the least credible of all agents examined (even when compared with a computer agent that expressed no nonverbal behaviour).

It is evident from the studies reviewed above that the use of nonverbal behaviours plays a crucial role in the development of positive user experiences with ECAs. Hence, it is important to understand whether the use of nonverbal channels influences the users' behaviour towards ECAs or the tasks they support. The studies reviewed in the following section attempt to explore this question.

4.2.2 The user's behaviour while interacting with the system

Investigating the effect of different methods for guiding the user's attention focus in multi-device presentations for public information systems, Kruppa and Aslan (2005) had their subjects watch parallel presentations on a system composed of a large stationary information system or personal mobile devices. In one condition, users were given a signal to switch their attention from one device to another by an ECA that could move from the stationary system to the mobile device and vice versa. In the second condition, the signal was given by an animated symbol, whereas in the third condition, the system did not warn the users at all. The researchers collected both objective and subjective measures of attention. The data from the objective measures did not favour any of the methods, but the subjective measures showed a clear preference for the ECA version. Based on these results, the researchers suggested that in order to allow a user to follow presentations spanning across mobile and stationary devices without putting too much stress on them, an ECA could be used as an effective method of guiding the user's attention focus. Other features of handheld ECAs include their ability to establish social bonds with users, as is demonstrated by the following study.

Bickmore and Mauer (2006) conducted an exploratory study on the impact of different user-agent interaction modalities on the ability of the agent to establish a social bond with the user. Participants were given four versions of an agent interface as follows: (FULL) a full version of the ECA (animation text and sound); (ANIM) an

animated ECA without nonverbal speech; (IMAGE) an ECA showing a static image of the character; and (TEXT) the interface without any character. Results revealed that, on social bond scores, there were significant differences between the animated versions of the agents and the other two modalities. Therefore, the researchers concluded that users establish stronger social bonds with ECAs that are embodied and animated compared with alternative modalities.

The studies discussed below have investigated the attention-capturing quality of ECAs in more detail:

An experiment by Witkowsky *et al.* (2001) used eye-tracking to measure the amount of attention drawn to ECAs. Participants interacted with an ECA (called James the butler) that informed them about different sorts of wine through both verbal and textual means. It was shown that an ECA attracts and holds the participants' visual attention in the interface, but also that this focus may have distracted from the products about which the agent was providing information. However, it could be argued that since participants devoted most of their time to reading the text appearing in a speech bubble above the agent it was hard for them to keep up with the agent's speech discourse, probably because of the unnatural, high speed voice generated by the text-to-speech system. If the agent's voice were more natural, there would have been no need for the bubble speech and, hence, a larger percentage of their attention would have been devoted to the products themselves. A similar study (Takeuchi and Naito, (1995)), on the attention-capturing quality of agents' utilizing gaze tracking reached the same conclusion. In particular, it was found that a character agent captures attention to a greater extent, than an animated arrow and that the participants in the latter condition needed more time to react than in the former. This extra time was interpreted as lack of concentration on the game by the paper's authors. One could argue, though, that the longer response times in the character agent condition may be the result of the additional cues (facial expressions, etc.) that participants had to account in making their decisions. In this way, the greater complexity of the situation may have led to increased response times.

Both of the studies above used ECAs that were not aware of where the user's gaze was directed. As demonstrated by Eichner *et al.* (2007) (see §4.1), a user-aware agent

using the appropriate interruption strategies can direct the user's attention effectively to the desired parts of the presentation. If this feature were possible in the above studies, the results may have been significantly different and possibly more favourable to the ECA interaction paradigm over the alternative ones.

Finally, the display configuration of ECAs seems to influence the users' behaviour towards an ECA system. The following study provides experimental evidence that supports this statement.

In an effort to understand the factors that influence users' behaviour with a virtual human, Johnsen *et al.* (2010) conducted an experimental study to compare the impact of two ECA display configurations on a number of important social dimensions. The first configuration presented the virtual human as life-size and was embedded in the environment and the other presented the virtual human using a typical desktop configuration. Participants in the first condition rated the system significantly better across all social dimensions (e.g., engaged, empathetic, etc.) than the participants in the desktop condition. It is worth noting that participants in the desktop condition focused more on the task at hand than on the social interaction with the ECA. Although, one would expect that the more engaging social experiences with ECAs would lead to improved task performances (see Bickmore's study in §4.1), the focus on the task may suggest that participants performed better in the desktop condition than in the mixed reality condition. However, since the study does not provide any performance measures, it is hardly possible to draw any conclusions.

4.3 Is there an ECA effect?

In the mobile space, there is evidence that a social ECA can impact positively on the task outcomes. A number of studies have also been reported with the aim of exploring additional effects of ECAs on the cognition of the mobile user (e.g., group learning and problem solving), but the outcomes of the evaluations were not released. While other variables, such as emotional intelligence, have failed to demonstrate significance in memory effects, they have produced superior subjective user experiences. This positive subjective effect has been reiterated by additional evidence about how users perceive an ECA with social characteristics (e.g., the desire to

continue working with it) more positively than those without social attributes (Bickmore (2007)). The indicative impact of ECAs on the user's trust towards the system has also been investigated, but failed to produce any significant evidence.

With regard to user behaviour while interacting with a system, evidence has been produced that users develop stronger social bonds with mobile interfaces that are embodied and animated over those that are not. Last, but not least, it was found that compared with non-embodied artefacts, an ECA is the most effective method of directing the user's attention focus on demand (Kruppa and Aslan (2005)). The issue of attention in ECAs was further investigated in ECAs for stationary computer applications.

The attention capturing quality of ECAs has been examined in more detail in immobile applications. The evidence suggests that a display with an ECA possesses an attention- capturing quality not possessed by other non-embodied displays. However, I argue that, in order for this attribute to be beneficial, an ECA must be able to resolve where the user is looking at all times. This way it can tailor and personalise its output to help the user to achieve his/her goal(s). A last piece of important evidence that impacts the user's behaviour is the display configuration of an ECA. The relevant study established that a life-size ECA can impact more positively the social perception of an ECA over a "typical" desktop configuration (Johnsen *et al.* (2010)).

In the case of a user's subjective impressions of a system, the evidence shows that an ECA with both nonverbal facial and body gesticulation can create positive user experiences compared with the absence of these communication channels. The use of these channels increases the usefulness and credibility of an ECA system. With regard to cognitive effects, the evidence supports the "Persona-Zero" effect (i.e., an ECA has neither positive nor negative effect on objective performance). However, as demonstrated by a more recent study (see Baylor *et al.* 2008), the use of gesticulation in an ECA can lead to improved user recall performance and a more positive learner attitude towards the agent. Hence, it is not the mere presence of a computer generated character (as in the "Persona-Zero" studies) that leads to improved task performance,

but rather a character is needed that uses proper verbal and nonverbal communication channels to amplify the effect of a message and intensify its meaning.

In summary, there is now some evidence that the use of ECAs on computer interfaces can actually benefit human-computer interaction in both mobile and immobile environments. Although some of the studies that I reviewed need to be revisited (e.g., Lim and Aylett, (2007)), and critics would argue that the produced effects are application-specific, nevertheless they are certainly a worthwhile starting point to inform proper designs for and ECAs for even more fine-grained evaluations. Furthermore, it is reasonable to assume that some of these effects may be universal. Specifically, as the user's perception of an ECA is affected by its display configuration, mobile devices with large displays and natural user interfaces (e.g., the Apple's iPad), should be preferred over more "traditional" mobile devices. Then, an ECA with nonverbal conversational behaviour should lead to higher knowledge gain under mobile environments as well. However, more research is needed to explore which aspects of the ECA's nonverbal behaviour (facial expressions and/or gesticulation) impacts learning and of what kind of information in these environments. In a series of exploratory studies (see §7.1 and §7.5 of Chapter 7), I attempt to explore the impact of an ECA with both facial expressions and gestures on the retention of information (of varying type and complexity), delivered during an interactive tour of an outdoor tourist attraction. A last effect that can be generalised is the effectiveness of user-aware ECAs in guiding the user's attention to the desired parts of the interface. I believe that such feature is of crucial importance, to the success of human-ECA interaction, especially in the domain of information presentation systems (mobile or stationary).

Finally, the effects discussed are likely to be domain-dependent, i.e., they are limited to the domain that the system is designed for. Other factors that moderate the complex relationship between the kind of interface used (ECA vs. not-ECA), and the user's subjective impressions and/or performance are the type of ECA used and the type of information the system provides. Getting a complete picture of how these factors interact is a difficult, if not impossible, task. Hence, there is a need for some form of ECA design standardization for both mobile and immobile environments. A standardised toolkit would remove any ECA/application dependencies in

experimental evaluation and would allow easier generalization and more meaningful comparison of results even between different domains. This thesis work, proposes the Talos toolkit (see Chapter 6) as a standard means for designing and evaluating ECA applications, primarily in the domain of interactive tour guides, but also in other domains as well (e.g., mobile learning applications, etc.)

4.4 Problems Left

As discussed above, a major step towards ECA exploratory studies (for mobile or immobile systems) with more brisk and clear-cut outcomes is the provision of a standardised toolkit for designing, developing and evaluating ECA applications. Such a toolkit would make it easier for researchers to create reservoirs of application-specific experimental data to enable better cross-domain comparisons. This would in turn, make it easier to draw more universal conclusions about the effects of ECAs on the user's attitudes and performance.

With regard to ECAs in mobile devices, it is evident that the existing literature is limited and further work is needed to build a consensus on their potential benefits on mobile interfaces. Narrowing my focus further to the studies of ECAs in mobile tour guides, I summarize below the limitations discussed in previous sections, and I add some additional observations that should be taken into account in future exploratory studies:

(1) There is sometimes a mismatch (for example, the Lim and Aylett study in §4.1) between the information the interactive guide provides and the actual physical environment. This distracts users from effectively relating the system's content to their surrounding environment.

(2) There is sometimes a limited use of multimodal[19] interaction. Recent evidence suggests that multimodal, embodied agents capable of user awareness and nonverbal conversational behaviour can, have a significant, positive impact on the user's impressions and performance.

[19] A multimodal interface reacts to verbal and nonverbal communication and generates appropriate output using verbal and nonverbal channels.

(3) None of the studies reviewed accounts for the effectiveness of the ECA in guiding the user's attention focus to specific parts of the physical artefacts/objects (a very important aspect of a real tour experience) about which the system is providing information.

(4) Although little is known about the effects of ECAs on human learning in mobile environments, research has been suggested (see Kruppa *et al.* 2005) that focuses on ECA interfaces that support collaborative learning in small groups. I argue that it should first be justified whether I should (with regard to the possible effects on learning) impersonate human characteristics when designing mobile interface agents with individual users before introducing them into group situations where the dynamics of learning are far more complex.

(5) While there is a near-to absence of evidence on the fundamental question whether the presence of an embodied conversational agent improves human-device interaction, studies have explored more fine-grained questions (e.g., emotional aspects of ECAs). I argue that some empirical evidence should first be established on this question before attempting to explore any additional aspects of ECAs.

Finally, an additional standardization problem common in both stationary and mobile empirical evaluations is the absence of a solid theoretical framework for supporting evaluations and interpreting the results. The direct consequence is findings that are often contradictory and equivocal among studies and are difficult to interpret. This, in turn, hampers the process of gathering requirements, measuring results and eventually refining the design. There is a strong need for a solid cognitive theory to start from and to drive empirical evaluations for both mobile and stationary setups, instead of some assumptions randomly chosen from fields, such as communication and psychology.

4.5 Experimental Questions and Working Hypotheses

Based on the discussions of the previous section, I therefore suggest that in order to rigorously and systematically investigate the use of embodied conversational agents in mobile interfaces; an empirical study in the area must address the following research questions: "Does the presence of a multimodal embodied conversational agent improve the accessibility and/or usability of mobile interfaces?" Given that some empirical evidence has been established by examining these questions, the following could also be investigated: "What aspects/attributes of a multimodal embodied conversational agent influence what aspects of accessibility and/or usability of mobile interfaces?" I have attempted to address these questions in the micro-domain of tour guide systems. In such environments, and in relation to the first question, my initial hypothesis was as follows:

H0: The presence of a multimodal ECA benefits the user's subjective experiences with a system but not his/her objective performance.

This hypothesis was tested in an early empirical evaluation (see Appendix C). Based on the results of this study, I derived a number of follow-on hypotheses focusing on performance with the aim of exploring how the presence of an ECA impacts specific aspects of the user's performance. The results of the three studies that were conducted with this aim are reported in Chapter 7.

For the second question my initial hypotheses were as follows:

H1: A multimodal ECA that uses well-synchronised non-verbal behaviours with speech positively impacts the user's performance and subjective experiences with the system.

H2: A multimodal ECA with the ability to perceive the user and react accordingly, positively impacts the user's performance and subjective experiences with the system.

In the same manner as above, I focused my original hypotheses on performance and expanded them with the aim of exploring how specific attributes of the ECA impacts

specific aspects of the user's performance. The results of the two studies that were conducted with this aim are reported in Chapter 8 (see experiments five and six).

4.6 Conclusion

A large amount of literature has contributed to evaluations of the impact of embodied conversational agents in stationary systems on the tasks that the users are asked to perform and on the quality of the user experience. These are usually on the screen of a desktop computer. However, when it comes to ECAs for mobile systems, the relevant literature is limited. In addition, there are some limiting aspects of the methodological approaches followed by these studies, as well as a lack of a sound theoretical framework for supporting the evaluations and interpretation of the results.

As an alternative, the research within this thesis has been based on a series of systematic and theoretically well-supported empirical studies to identify whether: a multimodal ECA can affect the user's experience, attitudes and performance with tasks in simulated mobile conditions; in order to uncover the aspects of those dimensions that were affected; and, distinguish factors that can be improved. By combining concrete measures and in-depth analysis of the results, the findings of these studies are presented as a list of design recommendations. A discussion of the methods and techniques that have been used in these experiments is presented in the next chapter.

Chapter 5 Research Methods

As discussed in the previous chapter, the experimental idea that the use of embodied conversational agents might improve human-device interaction was investigated with regard to various possible effects on the user's attitudes, experience and cognition. This chapter presents the research methods and techniques that were used to explore these effects, in the context of the chosen mobile domain - a series of mobile guide applications in an outdoor environment (i.e., the Monemvasia Castle[20]), a justification of their use, as well as their strengths and weaknesses in this research context.

In the first part of this chapter (§5.1), I outline the chosen styles of evaluation and their strengths and weaknesses, in the context of the current research. Following this, I present the chosen research methods and techniques, explain why I chose them, and discuss their strengths and weaknesses, in this particular context of use. In this part I also present a novel method for evaluating the cognitive accessibility of agent-based information presentation systems, using a combination of advanced evaluation techniques such as eye tracking and facial expression recognition.

In the second part of the chapter (§5.3), I suggest a comprehensive framework that encapsulates four key components of factors that would affect human users performing tasks with the help of ECAs in mobile environments. This framework draws on the relevant literature and theoretical models discussed in the previous chapter. It provides a number of crucial variables for each factor, amalgamated by theoretical principles, and a number of previous empirical works. Finally, given the constraints of the research (in both resources and time), I choose the most appropriate variables to test.

5.1 Styles of evaluation

Before I consider the chosen methods and techniques, I distinguish between two distinct evaluation styles that have been employed at different stages of the

[20] This castle is a popular tourist destination in the area of Peloponnesus, Greece with rich cultural heritage

evaluation: a pilot study that was performed at the requirements gathering stage in the field, that is the actual area of the castle of Monemvasia, and a series of exploratory studies performed under laboratory conditions across three countries (Greece, UK and USA)

5.1.1 Field Studies

A field study was used in the evaluation of an early prototype tour guide system (see Appendix C), in the actual castle of Monemvasia. I obtained measurements of users trying to find their way along specified routes and uncovering information about locations in the area of the castle with the help of the system. I chose a field study for the pilot, driven mainly by the need to generate results on the possible effects of ECAs that make sense for real-world usage. This means, that I was able to gather data that would have been difficult to obtain in a laboratory study, such as how the ECA impacts the participant's ability to physically locate landmarks, in order to navigate an assigned route, or how s/he interacts with physical objects for which the system is providing information. Finally, I wanted to investigate the practical problems and difficulties in conducting a field experiment in a busy tourist attraction such as the castle of Monemvasia.

To my disappointment, I encountered several practical problems that made me finally choose to simulate the mobile conditions in a series of lab experiments. My biggest challenge was the difficulty of recruiting volunteers from the visitors of the castle to participate in the experiments (ideally for free). It was very time-consuming to recruit participants at random from the visitors, as most of them were visiting the castle for a short stay or found the task too demanding to do for free. Another challenge was the difficulty of controlling the effect of the environmental extraneous variables. The light and noise levels, the weather conditions and the dense pedestrian traffic (from the incoming tourists) in the castle were often excessive. This, in combination with the interrupting questions from the passing pedestrians about the nature of the work, unfortunately made running the pilot studies a very unpleasant and difficult experience for the experimenter and the participants.

Perhaps the biggest challenge was when I tried to obtain the necessary permissions from the castle's archaeological authorities for conducting the next stage of the user evaluations (see Chapters 7 & 8) in the castle. The application was substantially delayed and eventually lost in the chaotic bureaucracy of the Greek state. As there was no foreseeable way to address these problems, I considered, as an alternative, the idea of simulating the environment of the castle in the laboratory, using high resolution panoramic images. My approach is discussed in detail below.

5.1.2 Laboratory Studies

This type of evaluation study was eventually used in the final stage of the evaluation, where the user experience of a series of prototypes was investigated with actual users under simulated mobile conditions. In particular, I asked each user to take a tour with the system in a quiet room, as if they were actually in the area of the castle of Monemvasia. Each route (landmarks and locations) that participants had to follow in the castle was represented by panoramic photographs and high-resolution video-clips projected on the wall by a video projection system. A small pilot was conducted prior to each experiment using a small group of users that tried to identify bugs with the software, problems with the data gathering tools and others. Participants were asked to think-aloud at all times with an experimenter present, whose responsibility was to record the session. The problems were normally corrected on site and the process was repeated once more. That way, I was able to correct any problems that could have disturbed the formal experimental evaluations with individual users.

However, the lack of context - for example, 'real' landmarks and locations, and realistic environmental conditions (e.g., no varying lighting conditions, no disturbing noises, phone calls etc.) - and the unnatural situation meant that I have recorded a situation that is a limited representation of the real-world. Although the immersive set-up created a feeling of being in the actual location and the imagery was made specifically to test navigation decisions (e.g., see §7.1), it was impossible to observe a user physically walk from one location to another and to make navigation decisions taking into account multiple variables (environmental or otherwise), as this is dependent upon the surrounding environment. For physical locomotion, a solution that will be examined in future experiments is the use of an electric treadmill that

would require participants to physically walk from location to location. Kjeldskov & Stage. (2004) followed a similar approach that could provide some valuable insights to help determine an effective experimental set-up.

5.2 Experimental Evaluation

A series of experimental evaluations was carried out (see Chapters 8 & 9), where six versions of a tour guide application featuring different ECAs were compared with appropriate control groups. I obtained quantitative, qualitative and usage data from different groups of users. The goal of the experiments was to provide some evidence to investigate the idea that the use of ECAs can improve human-device interaction and, in particular, with an interactive mobile guide application. To achieve this goal, a comprehensive evaluation framework was developed, which consisted of four key factors and various different variables within each factor (see §5.3), whose effect can be tested in empirical evaluations of such systems. Based on research and budget constraints I selected the most appropriate variables (see §5.3), to test in my experiments. In addition, a number of hypotheses were developed (see Chapter 8 & Chapter 9), to examine the possible impact of an ECA on the user's attitudes and performance. These were drawn from the initial hypotheses discussed in the previous chapter (see §4.5 in Chapter 4), and a pilot experiment conducted with an early prototype system (see the paper "*Humanoid Animated Agents in Mobile Applications: An initial user study and a framework for research*" in Appendix C)

There are of course some problems in using experimental evaluations in simulated mobile conditions in a laboratory. The first weakness is in the experimental setup itself: as participants in such a setting have to interact with two applications (i.e., the virtual guide and the castle imagery) at the same time, this may create synchronization and control problems. In order to create a realistic simulation, it is desirable that participants control the castle's imagery while standing. To achieve this, I asked them to use a Wii-Remote along with the mobile device in order to interact with the caste's imagery. The Wii-Remote was connected to the laptop running the castle's imagery, using Bluetooth. However, participants indicated that the particular setup was an uncomfortable experience as it was difficult having to synchronize their interactions with two interfaces while using two devices. Therefore, a simpler setup was followed

where participants interacted with the imagery, using a simple wireless mouse while sitting comfortably on a chair in front of the projector. Another weakness is the constant-presence of an experimenter in the room with the participant that may have made the participants feel uneasy. This observer effect may alter the way participants perform the specific task and provide biased data, or even worse, result in participants failing to complete the assigned tasks (Frank & Kaul 1978). I believe that I have minimised the influence of this effect by using the following methods prior the beginning of each experiment: a) the experimenter attempted to make participants feel as comfortable as possible, by introducing the system and the goals of the experiment, through a "casual" conversation with the user and b) designing an optional "introductory dialogue" for participants who either did not feel comfortable asking questions to the experimenter or who wanted to know more about the system and its functionalities.

I have chosen a number of appropriate techniques for collecting data in my experimental evaluations. Below, I elaborate these techniques separately, according to their use in the investigations of the potential effects of ECAs on the user's attitudes and behaviour.

5.2.1 Users' attitudes

Using an ECA, on a computer interface potentially influences the user's attitudes towards the system or the tasks it supports. A multi-technique approach consisting of three questionnaires and an interview was used to measure the participants' attitudes towards the cognitive accessibility and usability of the systems and the agents (embodied or otherwise). In addition, questionnaires were the technique of choice for measuring the mental effort required by the participants to complete tasks of varying difficulty with the systems, and their perception of the answers provided by an open natural language question-answering system.

5.2.1.1 Cognitive Accessibility, Usability, and Agents Questionnaires

The participants' attitudes towards the cognitive accessibility and usability of the systems and the agent were measured through the use of three questionnaires. All

three questionnaires support testing and identification of the user's subjective satisfaction with the systems and the agents and include various dimensions such as perceived intelligence, likeability and entertainment value (see tables D.1.7, D.1.8, and D.1.9 in Appendix D for a complete list of measures) (Adams 2005a; McBreen & Jack 2001; Catrambone *et al.* 2002; Adams 2005c). Participants were asked to indicate the level of their agreement with a questionnaire statement on a seven-point Likert (1932) scale, ranging from "strongly disagree" to "strongly agree". The seven-point scale has been shown to provide a more accurate measure of a participant's true evaluation than other scales (e.g., 5 point) (Finstad 2010). I have chosen questionnaires to assess how the users feel about using the systems and agents rather than a think aloud protocol in order to obtain quantitative comparisons of the effects of an ECA on the various aspects of the user's attitudes towards the system and of their views of various agents' qualities and attributes.

This approach has its pros and cons. Questionnaires place few mental and time demands on the user. This means that a participant would be able to answer them easily and relatively quickly, thus avoiding potential distress or discomfort (e.g., from a prolonged stay in the lab or from the complexity of the questions). On the other hand, questionnaires are prone to human error(s) (e.g., question(s) that participants ignored or indicated multiple preferences) and can become difficult to manage as the number of participants increases. However, these can be easily controlled if electronic questionnaires are used and administered on a laptop computer directly after each session with each of the systems.

When, using questionnaires there is always the issue of *validity* and *reliability*. Reliability is the consistency of the measurements, that is, whether the questionnaires measure the same way each time they are used, under the same conditions, using the same subjects. To test the reliability of the questionnaires I have measured their *internal consistency* by grouping questions that measure the same concept (e.g., ease of use) and correlating between each questionnaire response from a group to determine if the questionnaire reliably measures the specific concept (see Appendix D and Appendix E for more details). The validity of a questionnaire is the degree in which each set of questions measure what it is intended to measure. It is ensured by a solid theoretical framework (see Chapter 3), and a strong relevant literature base

(Adams 2005c; McBreen & Jack 2001; Catrambone *et al.* 2002; Adams 2005a; Adams 2006). Finally, although questionnaires typically generate valid results, a number of factors may affect the user's satisfaction with the system or the agents (e.g., give rise to unpredictable high or low levels of satisfaction) which the experimenter did not anticipate in the questionnaire. For this reason, following the questionnaires, a contextual interview has been performed that has provided each participant with an opportunity to expand on any underlying issues.

5.2.1.2 Post-task Interviews

Post-task interviews were used as a supplementary technique (administrated immediately after the questionnaires) for eliciting more information about the user's impressions of the cognitive accessibility and usability of the systems, the agents and to help disambiguate the participants' responses in the questionnaires. In addition, interviews have served as an effective mechanism for revealing any particular troubling points in the design of the systems and/or the agents. The interviews followed a semi-structured form, i.e., they facilitated an open discussion with the user, but they were guided by a general written outline of the topics to be covered and the information required. I felt that this form was suitable, because, given the dynamic situation of the user, it helped focus the purpose of the interview on the issue at hand, but it was also flexible enough, to allow participants to introduce and discuss any issues which they found relevant. On a more important level, a semi-structured approach enabled me to tailor the duration of the interview to match the time constraints of each participant.

To analyse the interview data, I followed a custom-made approach. I used this approach to analyse the interview data in experiments one, two, three, five and six. In particular:

I divided the participants of each experiment into a "Feedback" and a "Confirmation" group. Each group consisted of equal number of participants randomly chosen from the experimental conditions. If a participant had not provided feedback, s/he was excluded from the groups. I looked in the feedback group for comments based on:

- Frequency: These were groups of comments that frequently arose around a specific event encountered or feature of the systems (e.g. ECA design, experiences with the multimodal content, etc.). As a frequency threshold for these patterns, I defined 40% of the total number of participants in the feedback group.

- Fundamentality: These were comments that although did not frequently arise were deemed by me to be fundamentally important in terms of the possible effects of ECA on the user's experience of the prototypes.

Then, I looked in the confirmation group for comments that corroborated the patterns and/or comments of the feedback group. If a match was found in the confirmation group, the pattern/comment of the feedback group was considered as corroborated. If a match was not found in the confirmation group, the pattern/comment of the feedback group was considered as uncorroborated. As participants in both groups were taken at random from the experimental conditions, views were mixed and therefore uncorroborated patterns/comments could arise in the analysis. Only the patterns or comments that were corroborated were used in drawing conclusions from the experiments.

The one-way flow of information in the interviews though, means that I could gather information that merely represented the participants' perceptions of the system. A focus group with a small number of participants could also have been conducted to examine in greater depth the participants' perceptions and opinions about the systems. The difficulty of setting up such a group meeting however, even in lab conditions (e.g., getting the group members together at a specified time), made the use of focus groups difficult in the current research.

5.2.1.3 Cognitive Workload Questionnaires

A set of seven-point Likert (1932) questionnaires (disagree-agree), was used to measure the participants' subjective cognitive workload (see §5.2.3.4 for a description of this measure) when navigating routes of different complexity and uncovering

information, with variable degrees of difficulty, from locations in the area of the castle. These questionnaires were based on the Simplex-Two theory (see §3.2.2 of Chapter 3) and addressed dimensions such as complexity of navigation, user frustration, complexity of the presented information, etc. I decided on the sole use of questionnaires to assess aspects of the user's mental effort, rather than psychophysiological techniques (e.g., eye-tracking) for two practical reasons: a) part of the formal experimental studies were conducted in Greece, which made it difficult to borrow and move such expensive equipment abroad and b) the unavailability of such equipment even for internal university use.

5.2.1.4 Q&A Questionnaire

A 10 point questionnaire (1-10 scale with 10 being a perfect answer) was used to measure the participants' subjective impressions of the answers provided by a natural language Question-Answer (Q&A) system designed specifically for the castle of Monemvasia. The questionnaire was used during the test and addressed dimensions such as clarity, sense, accuracy of answers, etc. Subjective impressions could also have been collected using post experimental interviews. However, any qualitative data collected after a session with the system would have been incomplete, as the impact of an answer is temporary and it should be measured once the system provides it. In addition, it is easier to correlate quantitative measures of the users' perception of the answers with performance data generated from the retention tests administered immediately after a session with each of the QA systems.

5.2.2 Users' behaviour

The use of an ECA could change the user's behaviour in his/her interactions with a system which could improve the outcome of the interaction process. For instance, the user might want to spend more time on a system with an ECA than a system without it, which could result in a more extensive knowledge of the system's domain. A questionnaire was used to measure aspects of the user's attention to the task with the ECA. In addition, I employed a multi-technique approach that consisted of device logs, direct observation and written tests to measure the users' performance with the systems. Then, a technique that combines eye tracking and facial expression

recognition provided further insights in the perceived cognitive accessibility of the content presented by the systems. This technique significantly augmented the data captured with the more "traditional" techniques, such as written tests or questionnaires.

5.2.2.1 Attention-Specific Questions

A number of yes/no questions were included in the cognitive accessibility questionnaire (see table D.1.7 in Appendix D), to measure the effectiveness of the ECA in guiding the user's attention focus towards physical objects in his/her environment. I used questions to explore this aspect of the user's attention in all of my experiments for two reasons: first, it was difficult to get access to an eye tracker, let alone to move it to Greece and set-it up without the help of an expert. Second, when an eye tracker became available to me, I used yes/no questions in combination with data from eye-tracking, to examine any differences between the objects participants said they saw in the questionnaires and those they actually casted their gaze upon the interface (see §*Eye tracking – Gaze trails* in Chapter 8).

5.2.2.2 Protocol Analysis

During the experiments, computer logging was used, to automatically collect measurements of the user's total time to complete a tour with the system, and the number of questions asked on each of the locations in a route. Although device logs could also have been used to automatically measure the number of user's navigational errors (see §5.2.3.1, §5.2.3.2 and §5.2.3.3 for a description of these measures), the simulated environmental conditions made their use unnecessary. In particular, more than one button was placed on the panoramic applications at points where users had to make important navigation decisions. A navigation error was defined as participants clicking on the wrong button on the screen, which activates an error message window. This process made the collection of navigation errors as simple as observing the users and recording any navigational errors on a piece of paper. Feedback was provided in erroneous situations to help participants overcome these errors. However, because of the unique nature of the panoramic application (360 degrees viewable images), users may simply fail to see the correct button on the screen (the ones that load the next step

of a route) and hence, fail to click it. However, the impact of this effect may be minimized if participants are given time to become familiar with the panoramas (and their use) at the beginning of each experiment. In addition, as already discussed (see §5.2), the mere fact of being observed can alter the way participants navigate the assigned routes and therefore produce biased performance data. But, by allowing some time for the participants to become more inured to the experimenter's presence, the influence of this effect could be minimized.

5.2.2.3 Written Tests

In addition to the above performance measures, I have measured the effect of an ECA on the participants' retention of the information presented by the systems, via post-written tests. The tests were administrated immediately after participants completed a session with each system. Each participant was asked to indicate his/her knowledge of the information presented by the systems on an electronic short-answer format test (see Appendix D for one of the retention tests used in the experiments). I chose this particular style of testing rather than other formats (e.g., open-ended questions), because it does not make excessive demands on the user in terms of complexity and time of completion – some very important requirements for the participant of my experiments who is already loaded with enough tasks – and can also be analysed more rigorously.

5.2.2.4 Eye tracking and facial expressions

To get an accurate picture of the perceived cognitive accessibility of an agent-based information system one needs to measure the user's immediate experience while the task occurs. "Traditional" techniques such as questionnaires capture the users' experience after the task is complete and are generally prone to various confounding variables (e.g., human errors in administrating and scoring, questions users failed to understand correctly, etc.,). For these reasons, a technique that combines data from eye-tracking and face expression recognition was crafted to collect behavioural measurements while the interaction occurs. The human face is one of the strongest indicators of a person's cognitive state and hence how s/he perceives stimuli (information content, images, etc.). If data from both facial expression and eye-

tracking are combined and analysed carefully they may reveal valuable insights into various aspects of the cognitive accessibility of the content. Facial expressions can reveal the emotional state of the user when encountering the information content (e.g., which part requires the users to think more intensively or which part is more confusing, etc.). Then, data from eye-tracking can reveal which part of the interface the user was looking at when the particular expression(s) occurred. These data can also help explain, for instance, why a user scored badly in retention test or why s/he rated the likeability of the system high. Although someone could argue that it is easy for users to conceal their emotions by masking their facial expressions, Wang and Marsela (Wang and Marsella 2006) observed face expressions indicating a range of emotions (e.g., boredom and anger) in users of a dungeon role-playing game with embodied virtual agents. Last, but not least, measuring brain activity (Simple Usability[21], 2013) captures emotional engagement more accurately than my proposed method; however my approach is far cheaper and, most importantly, it is invisible to the user. Measuring brain activity requires the use of sensors attached to the head of the user. This will most likely create a highly unpleasant and intrusive experience for the majority of the users.

5.2.3 Some evaluation measures

This section gives a detailed overview of the quantitative and qualitative data that were collected during the empirical evaluations of the systems. The data were collected using a mixture of device logs, direct observation, questionnaires, eye tracking and face expression analysis.

5.2.3.1 Timings

The time taken to complete a tour with the aid of the systems has been used as an indicator of the user's performance with the systems. In particular, the user's start and end times to complete the set of the assigned tasks, i.e., to navigate his/her way along a specified route and to uncover information about particular locations along the route, was recorded as an indication of his/her time performance. Timings were

[21] http://www.simpleusability.com

automatically measured by enabling the device to record them in interaction logs. Although recording performance times in a simulated environment has little meaning (as users do not physically walk around an attraction), nonetheless this measure can provide valuable user insights, such as the degree of user engagement with each of the systems, the quality of question-answering dialogues (Q&A), etc.

5.2.3.2 Number of Questions

The number of questions asked in a question-answering session with the systems was used as an indicator of the degree of participant engagement during the presentation of information about a particular location of a route. In particular, the number of questions following a presentation about location X could reflect the participants' engagement with the system during this activity. One can suggest that the more questions a participant asks, the more engaged s/he is with the system and, hence, more eager to explore the available content in more depth (something which could also have an impact on the time s/he spends to complete a tour). Again, the total number of queries asked per participant in each location of a route was identified from the device interaction logs.

5.2.3.3 Errors

The number, or rate, of navigation errors was used as another indicator of the user's performance and success in using the systems to navigate the assigned routes in the area of the castle. Evaluating the impact of the systems on the ability of users to take navigation decisions was a difficult task in a simulated environment. On the panoramic scenes which simulated the junctions where the user needed to decide where to go, I placed more than one button. If the user clicked on the correct button, the next panoramic scene loads, if not an error message window appears warning the user that this is not the right way to go. Such errors were noted by the experimenter on a notebook. A user was defined as "lost" if s/he clicked on the wrong button (and hence, would have deviated from the planned route) and the experimenter had to intervene in order to get him/her back onto the route.

5.2.3.4 Perceived Workload

Cognitive workload relates to the mental effort required to perform a task. In this scenario, this refers to the cognitive resources required by a user to navigate an assigned route and to uncover information about a number of locations in the area of the castle. The user's perceived workload was used as an indicator of his/her satisfaction in completing these tasks with the aid of the system. The workload was measured through a seven-point Likert-scale (1932) questionnaire and included various measures of cognitive difficulty, such as complexity of navigation, frustration, complexity of the presented information, etc.

5.2.3.5 Gaze Trails

A gaze trail showed the order in which participants cast their eyes over the interface and it was produced automatically by an eye-tracker. It also showed for how long they looked at each section. This information can be of particular importance to agent-based systems as it can reveal whether users looked at the ECA, and which elements of the avatar attracted attention the most (e.g., hair, body, etc.). There is an abundance of mixed evidence in the relevant literature that avatars attract attention that could result in superior or inferior performances. However, the issue of attention has never been thoroughly investigated with the use of an eye-tracker. It is an open question, for example, over what attracts the attention the most to an ECA and how this is correlated to a potentially good or bad performance.

5.2.3.6 Heat Maps

In addition to gaze trails, which produce eye tracking data for individual participants, a heat map produces an amalgamation where all participants looked and for how long. The "hotter" the area, the more it was looked at and noticed by the participants. This can provide an overall picture of which part of the interface was looked more by the users, and how this overall behaviour relates to the overall performance of each group or individual participant.

5.2.3.7 Facial Data

It is highly likely that the information content of a system induces a range of emotions in the user (e.g., boredom, sadness, etc.). These user emotions are most likely to be communicated through their relevant face expressions. Facial expressions were recorded using a camera attached to the mobile device and later analysed using a mixture of manual and automated analysis (e.g., the SHORE engine (Fraunhofer Institute 2010)). The user's facial expression(s) evoked by each part of the content were correlated with gaze trails, heat maps and performance data. Furthermore, the group's facial expressions were correlated with heat maps and the group's overall retention performance.

5.3 A Framework for research on Embodied Conversational Agents (ECAs) for mobile guide interfaces

In order to effectively investigate the use of ECAs in mobile guide interfaces, I need to consider the key factors that influence the cognitive accessibility and/or usability of such interfaces. For this reason, I proposed a comprehensive framework for studying ECAs in mobile interfaces, which consisted of four key components: differences in users such as computer-experience, age or gender; aspects of how the agent looks, sounds or behaves; different mobile environments; and different task features. This framework, draws on a number of previous frameworks (Xiao *et al.* 2002; Catrambone *et al.* 2002), and theories discussed in the previous chapter. Below, I discuss a number of variables that are possible within each factor and conclude with those, that I am interested to test, as part of the current research programme.

Factor 1: Features of the User

Potential users vary, of course, in many ways. However, based on aspects of the Simplex-Two and Distributed Cognition theories (discussed in §3.2.2 of Chapter 3), I can derive certain features, that may be quite likely to affect how accessible and/or usable a user finds an interface with an ECA. These features include:

Perceptual capability: The ability of a user to perceive the information provided by an ECA in a mobile setting may vary greatly across individuals. For example, a user with a hearing impairment may find it difficult to understand the navigation instructions provided by the system if the ECA's non-verbal behaviours (e.g., lip-synchronization, pointing gestures, etc.) are not constructed based on his/her special requirements. Then, the different perceptual capabilities of users with regard to the use of system symbols and the non-verbal behaviours of the ECA may also need to be considered. For instance, it is a well-known fact that certain non-verbal behaviours are perceived differently by users of different cultural backgrounds.

Cognitive capability: The user's cognitive capabilities in terms of memory capacity, information processing, problem-solving skills, etc., may vary greatly among the users. A user who can readily acquire new tasks and implement them with relative ease would most likely prefer an ECA that is reactive and provides information only when directly addressed, whereas a user whose such ability is limited would require a more proactive ECA, that could provide him with more feedback in order to perform his/her task sufficiently well.

Age: The above factors can be differentiated greatly among users of different age groups. The users of older age groups often have reduced perceptual and cognitive abilities as well as motor responses compared with those in the younger age groups. Therefore, it is reasonable to expect that the attitudes and performance when interacting with an ECA also vary across different age groups. For instance, one might hypothesize that an ECA that is not designed with the needs of older adults in mind distracts from the successful completion of a task rather than be of support towards that direction.

Gender: The user's gender can play a role in the overall perception of the agent's characteristics. For instance, McBreen and Jack (2001) showed that for an electronic commerce desktop application, female participants deemed a set of female cartoon-like agents to be more polite than the corresponding males, and male participants thought the male cartoon-like agents to be more polite than the female ones.

Other variables: Other user-related variables include background knowledge, culture (a variable already discussed in §3.3.2 of Chapter 3) and device experience.

User data from the relevant ECA literature is highly sporadic. There has been no attempt so far to aggregate any user behavioural data into usable personas. Personas represent groups of users and they can enable the easy sharing of user behavioural data in the academic community. In this research, I synthesised the data generated from six experimental evaluations into highly usable personas, representing groups of users across three countries (Greece, UK and USA). I have made this data available to ECA research communities with the aim of becoming useful to other researchers.

Last, but not least, if I look into the possibility of ECA interfaces that support group work and social interaction between people, the variables discussed above would still apply, but would need to be adapted to take into consideration the requirements of user groups instead of individuals.

Factor 2: Features of the Agent

Like users, ECAs can vary in a wide variety across several features. These features include:

Visual presence: There is limited empirical evidence to support the argument that an agent should visually appear in interactive mobile guide interfaces. Due to the scarcity of prior research in the area, it has not yet been investigated whether anthropomorphic agents actually enhance mobile guide interfaces or not. For example, it could be found that voice output alone is sufficient for the interaction to take place successfully and for the task to be completed well.

Attention-grabbing abilities: A very important component in the agent's multimodal abilities may be that of sensing the user and reacting appropriately. Although the use of computer vision technologies have endowed ECAs with the ability to sense various user features (e.g., emotions, head position, etc.), it is how an ECA reacts to multimodal inputs that can impact the user's perception and performance with the system. For example, if a user is "bored" during a presentation

about a location of a tour, humour can be used as an effective strategy to get his/her attention back whilst not affecting the flow of the presentation. On the contrary, a too-forceful strategy probably irritates users and seriously disrupts the flow of the presentation.

Amount of embodiment: With respect to amount of embodiment, a number of values are possible (e.g., "head and shoulders" and "full-body"). An early theoretical study in anthropomorphic characters for mobile devices has suggested facial agents as a proper amount of embodiment (Cowell *et al.* 2003). This idea, however was not based on any empirical investigations of actual mobile guide applications and, therefore, it cannot be accepted as conclusive. Only proper empirical studies of functional mobile applications can provide an answer to the question of embodiment.

Types of nonverbal cues: The following behavioural nonverbal cues are possible, including: facial expression, eye contact, paralanguage gestures, and posture. Cowell and Stanney (2003) demonstrated that the portrayal of credible nonverbal behaviours plays an important role in the user's perception of trustworthiness of the system (and, thereby, of the information provided by the system - see §3.3 of Chapter 3 for a more detailed description of their work). However, as their work was focused on a desktop application, I cannot generalise their findings to mobile interfaces. Careful empirical work is needed in this area that would also account for the perception of these nonverbal behaviours from users of different social contexts.

Competence: Similar to the effect of the type of non-verbal cues is the effect of competence. I hypothesise that in the context of mobile guides, an ECA that uses minimal non-verbal cues would be perceived by users as less competent for the job it was assigned to do compared with an ECA that uses a range of non-verbal gestures of communication. If proper verbal cues were added to an ECA (e.g., the ability to pause between the sentences of a presentation) then the competence effect should become even stronger. Similar to the real life guidance scenario, users most likely find a human guide that speaks continuously and displays minimal non-verbal cues to not be suitable for the task of visitors' guidance, which requires the sensitive use of both verbal and non-verbal channels.

Natural language capabilities: The ability of an ECA to process natural language and respond appropriately is of major importance to ensure effective question-answering (Q&A) dialogues. Script-based approaches are generally very robust in matching an input to an answer in the database. However, as scripts do not process the meaning of human language, the dialogue designer must craft all possible options for a single question to match the best answer in a database. An alternative is to process the semantics of human language, which, however, requires the input to be formed in the correct grammatical form. This is most likely to result in users having to rephrase the same question several times. Although this could benefit performance (as the user has to review the content a number of times to rephrase the question), it may be highly detrimental to the overall user's experience.

Modality of communication: It is an open question how users should communicate with an ECA in mobile environments. Given the complex technical challenges involved in creating an ECA capable of accepting speech input and the limited ability of users to enter text using a keyboard under mobile conditions, one can assume that the use of menus with text phrases would be one of the most appropriate methods for communicating with an ECA. However, if the technical problems could be solved, a full scale empirical evaluation may be made to evaluate the above three types of communication.

Other variables: Other agent-related variables are "gender", "ethnicity", "age", and "personality"

Factor 3: Mobile Computing Environment:

The environment in which the mobile guide operates can of course, vary in many ways. These include:

Type: I distinguish two types of mobile environments (of course, other types of mobile environments such as that of a car are possible) for which a mobile guide system and an ECA must be tailored: an indoor mobile environment (e.g., a museum or an art exhibition) and an outdoor mobile environment (e.g., a castle, ruins of an ancient temple, etc.). Because of the unique nature of each environment type, the

potential effects of ECAs on the user are also distinctive. For example, in an outdoor environment the task of uncovering information about an attraction may be, by itself, a less efficient task than in an indoor environment. This is because the user may have to devote more mental and physical resources to completing the assigned task. For this reason, it cannot be assumed that the existence of an ECA results in the same effect on the user's learning across the two environment types.

Characteristics: The characteristics of the environment in which the user is located (either indoors or outdoors) can vary in a variety of ways. These characteristics could affect the design of the mobile guide system and, hence, the potential effects of an ECA on the user. For instance, in some outdoor environments, dense distribution of buildings could make the use of location sensing equipment (e.g., a GPS device) a very ineffective method of navigation. The different methods that could be used (e.g., landmarks) may have different kind of effects on the user (and thereby affect differently the user's perception and performance with the system).

Factor 4: Features of the Task

The tasks in which the user is asked to perform with the aid of the agent can also vary in many different ways. Some of these features are:

Navigational complexity: The complexity of the navigation task might be simple, in which the user has to find easy-to-find locations in an area. Alternatively, the user might be carrying out a more complex task where s/he has to locate more hard-to-find sites in an area. An ECA could have different impacts on the perceived workload and performance of the users with the tasks of different complexity levels.

Information personalization: An important feature of mobile applications is the ability to provide information, tailored to the location and profile of the user. Of course different levels of personalized information (e.g., according to the user's background knowledge) may have a different impact on the perception and performance with the ECA. Although, more research is needed to define the appropriate levels and values for this variable, a systematic exploration could be

started by considering the simple level of information difficulty with values simple and technical information.

In my studies, given the limited availability of technical and human resources and time for the completion of the research, I addressed the following seven variables: user gender, agent's visual presence, competence, natural language abilities, attention grabbing abilities, task navigation complexity, and task information difficulty.

5.4 Conclusion

The methods and techniques to be used in the current research were discussed in this chapter. This discussion was not intended to be an extensive treatment of all possible methods, but rather a justification of the most suitable ones for my needs. Clearly, in future work, other methods can be explored, but at the moment they do not appear promising. However, before applying the chosen methods and techniques in an actual experimental situation, there are two issues that need to be considered carefully. These are, firstly, to identify the variables that should be tested and manipulated in the context of the current research and secondly, to validate the appropriateness of the chosen methods in the field of study (and thereby refine them as necessary).

In order to address the first issue, a comprehensive framework was developed, composed of a number of factors and variables that should be considered, when evaluating the effects of ECAs on the user of a mobile guide application. However, not all variables can be tested in the current research and hence, the most appropriate ones are chosen. With regard to the second issue, a pilot experimental approach was adopted prior all formal experimental evaluations to refine the selected methods and techniques. In addition, the combined eye-tracking and face expression recognition method was validated in a formal experimental evaluation.

In the next chapter, the design and evaluation of the Talos toolkit – a toolkit to aid the rapid prototyping of mobile guide applications with ECAs - is discussed, along with several technical challenges that I encountered during the design process.

Chapter 6 Talos Toolkit

Talos Toolkit Design and Development

This chapter introduces the Talos toolkit, a novel open source authoring toolkit to build and evaluate mobile applications with embodied guide agents. Even though building the UI toolkit is not the primary focus of this work, a number of its components were partially implemented with the aim to build the prototype applications I used in my empirical evaluations (see Chapter 7 and Chapter 8). In order to motivate and guide any interested researchers or developers in the full implementation of the toolkit, its core information architecture is presented along with a number of heuristics that should guide its actual UI design.

To set the scene this chapter first sets forth (§6.1) the motivation of this development work. Then it discusses a cognitive walkthrough of two existing toolkits (i.e., the ICT Virtual Human (ICT, 2010) and the Guile3D toolkits (Guile3D, 2010)) with the goal of producing a number of heuristics for a usable UI design of Talos.

Finally the key requirements for Talos are presented and follow this with a full description of its information architecture and key components. This chapter continues with the presentation of the design, development and evaluation of the natural language component of the toolkit. In addition, as a result of the evaluation, a number of improvements are suggested to be implemented in the component. The discussion is presented in the second part of this chapter (§6.4-§6.5).

6.1 Motivation

Talos is an integrated toolkit that streamlines the development of fully multimodal ECAs for mobile guide applications. The toolkit focuses on the integration of various technologies into a single platform to provide a cheap and simple-to-use solution for researchers and application developers to build and evaluate ECA prototypes with users. Until recently, there were no integrated development tools available to the ECA community. The approach taken in most of the existing work was to either build the

various software modules needed from scratch and/or buy any other required components. However, given the high-degree of complexity inherent in the development of ECAs and the high cost of any market components, the lack of tools required developers to either build very superficial ECAs or to avoid developing ECAs altogether. The launch of the ICT (ICT, 2010) and Guile3D (Guile3D Studio, 2010) toolkits has a high potential to change that and make the domain more accessible to researchers and application developers. Nonetheless these tools are either exceptionally complex to use (i.e., the ICT toolkit) or closed-source (i.e., the Guile3D toolkit) that prevents modification of the toolkit in any way or the addition of any new components. In addition, both toolkits have been designed specifically to utilize the hardware resources available in stationary systems. For example, although the Guile3D character engine is compatible with most recent mobile devices, the components needed to add multimodal functions require the use of high-end hardware. Furthermore, none of these tools provide integrated methods for building the scientific instruments needed for the evaluation of systems with users.

6.2 Evaluating the ICT and Guile 3D toolkits

In this section, I perform a cognitive walkthrough of the ICT Virtual Human and Guile3D toolkits. To the best of my knowledge, these toolkits are currently the only complete toolkits for ECA authoring and development available. I aim to identify any usability problems with the toolkits and use what I learn to generate a list of actionable recommendations to guide designers in a usable UI implementation of Talos.

6.2.1 The ICT Virtual Human Toolkit

The ICT Virtual Human toolkit (ICT, 2010) is a collection of modules, tools and libraries that facilitate the creation of ECAs. The toolkit includes modules for natural language interaction, nonverbal behaviour and visual recognition.

Three tasks are examined:

1. To create a Question – Answering (Q&A) Dialogue
2. To add ECA gesticulation to the responses

3. To add multimodal input

The first task is about creating a question-answering dialogue in a certain domain. For the purpose of the evaluation, the domain of the prototypes was selected, i.e., a tour of a medieval castle. Within this domain, Brad (see figure 6.1) provides information about an attraction at the castle and answers a range of questions about the attraction upon completion.

Figure 6.1: Brad the default ECA of the ICT toolkit (source: ICT, 2010)

The second task is about adding ECA gestures and correct synchronization with the responses. The third task requires the expert to add multimodal input (i.e., head positions, head gestures, eye gaze and eye gesture) and appropriate reactions to the character. To perform the walkthrough, a list of the actions to be undertaken to complete the tasks discussed above, with the interface is required. To conform to the cognitive walkthrough methodology, four questions were considered for each major action completed (Travis 2010, Adams 2005b). The complete cognitive walkthrough is presented in Appendix A.1.

1. Will the users realistically try this action? Would the action occur to the user to do?
2. Will the users perceive the control for the action? Is the control visible?

3. Once users find the control, will they recognize that it is the one they want to complete the action?
4. Once the action has been taken, is the feedback to the users appropriate, so they can go to the next action with confidence?

It is assumed that the users of the toolkit are not novices in the area of ECAs, but that they are researchers interested in the domain and with some basic programming skills. These users would want to modify the Brad character provided to conduct formal evaluations in specific domains. Last, but not least, those users have reviewed the basic documentation and tutorials provided with the toolkit.

Although it was not possible to evaluate the full toolkit (as some of the components are command based), the cognitive walkthrough uncovered a number of usability problems (see Appendix A.1). To summarize:

First, the toolkit makes creating a new GUI character an unnecessarily complicated and time consuming process. It requires the user to go through a labyrinth of options where s/he has to disambiguate jargon terms almost at every step of the process. A more reasonable approach would be to enable designers to:

1) Create a character and adjust its properties/settings from the same panel of the editor. This will make it easier for designers to link characters to their associated properties in the system.
2) Create characters that have a default connection to the rest of the toolkit modules.
3) Define the states of the dialogue as a character property (e.g., “states”).
4) Set the initial state of the dialogue from within the above property.
5) Access important character properties and their background code by default. For example, the “Type” property used to handle off-topic responses, and a property needed to receive messages from the computer vision module, should be made available with the toolkit installation.

Second, the process by which the system learns the question-answer mapping is perhaps the biggest problem of the toolkit. It uses a statistical text classifier for

mapping questions to question-answer pairs in the database. Designers can tune several of the classifier parameters, assuming of course they have the necessary knowledge to do so. The training process should take place in real-time, during the entering of the question-answer pairs in the database. Unless explicitly requested by the user, such advanced functions should be hidden and fully automated.

6.2.2 The Guile3D Toolkit

To the best of my knowledge, currently the Guile3D toolkit is the only commercial solution currently available on the market. It has been designed to automate daily computer tasks (e.g., searching the web, checking email, exploring multimedia files, etc.), and it consists of a number of modules, some common (e.g. natural language communication) and others more advanced (e.g., the Smart Home module).

The complete cognitive walkthrough is presented in Appendix A.1. Although this toolkit is more mature than the ICT Virtual Human toolkit, my evaluation uncovered a number of usability problems:

First, based on experience with other interfaces I would expect the toolkit to save a Q&A database in a format directly readable by its dialogue and animation engine. Currently, this task requires two steps ("Save" the database and "Compile" the database in a readable format by the engine) that could lead even advanced developers to miss the compile step.

Figure 6.2: Denise the virtual human assistant of the Guile3D toolkit (source: Guile3D, 2010)

Second, although there is "click and insert" support for custom control tags[22], a) it is not clear what these tags actually do, and b) it is unclear which of these tags are designed to control Denise's facial expressions or other animations. A more reasonable approach would be to enable designers to:

1) Save a Q&A database into a format directly readable by the toolkit's dialogue and animation engine.
2) Select tags to insert to answers from categories that clearly indicate their function (e.g., animation, email, etc.).
3) Get appropriate feedback (ideally visual) that disambiguates the function of each tag.

6.2.3 Design Heuristics

The lessons learned from the evaluation of the toolkits were integrated with my own "expert" recommendations and the result was compiled into a list of general principles for a usable UI implementation of Talos:

Q&A Authoring:

1) Visual organization of Q&A nodes should be aided by automatically clustering nodes of the same information space (e.g., topics, subtopics etc.). This would enable a much easier search, modification and manipulation of the system's content.
2) Group variations and root questions in nodes and offer visual access to easily edit their content.
3) Enable direct control over the properties of each node (e.g., topic(s) to which it belongs, animation tags, etc.). The most important node properties should be made available by default, but the toolkit should allow the development of custom-made properties using an easy-to-use scripting language.

[22] Control tags are XML tags designed to control Denise's various features and functions.

4) Enable auto-saving of knowledge-bases in a format directly readable by the toolkit's animation and dialogue components. Inform the user about the result of the process through appropriate feedback at a reasonable time.

Avatar Animation:

5) Categorize animation tags according to function (e.g., pointing, beat gestures, etc.)
6) Use a visual tool to enable developers to modify the rules needed for automatic tag annotation of input texts.
7) It is important to seamlessly blend the character's spoken text with its accompanying face expressions and/or gestures.

Multimodal input:

8) Provide support for users to access the full range of functions supported by each of the vision components of the toolkit. For example, autocomplete could be used when creating emotion rules for the input modalities module.
9) Enable the authoring of character reactions to input stimuli without the need to write any code. Simple English-like commands should be used.

Toolkit as a whole:

10) Enable communication between the modules of the toolkit (and any other new module) without the need to be explicitly requested by the user. The status of the connection should be visible to the user through proper interface feedback.
11) Avoid loading each module of the toolkit all at once. Make each module accessible only on a user's request (e.g., through natural language commands or a module menu)

Although the list is far from comprehensive and definitive, I hope that the potential developers of Talos will find it useful. In order to automate the most tedious tasks in the development of the prototypes, some of the toolkit's modules were prototyped in

some form. Below, I discuss the most significant developments. For the rest of the components, see Appendix A.2.

6.3 Design Requirements

In this section, I explore the core design requirements inherent in the architecture of Talos. In order to build a toolkit that meets the requirements of researchers and application developers, I first sought to understand their needs in the particular domain. This process was mainly informed by a significant "expert" knowledge gathered from the design, development and evaluation of several ECAs for mobile applications. The requirements list generated from this investigation ultimately determined the functionality and feature set of Talos. Here, by discussing in detail the requirements, I justify the design decisions I made in the development of Talos.

6.3.1 Universal Compatibility

The full multimodal behaviour of ECAs should be available independently of mobile operating systems and device hardware specifications. This means that the character engine and any other components required for multimodal input and output should be platform-and device-independent. In the current state of technology, achieving Universal Compatibility (UC) is difficult. Although the components for multimodal communication can be web-enabled and become platform-and device-independent, the Web has no inherent support for rendering the complex 3D graphics required for the realization of full ECA behaviours. The only alternative currently is the Flash platform[23] (Adobe, 2013), which, however, enables ECAs of very limited realism (e.g., Bickmore *et al.* 2009). Therefore, Talos was required to provide a stand-alone mobile player that featured stable network communication, along with its source code and complete documentation. This way, developers can port it to different platforms with minimal effort and in this sense; achieve the goal of "Universal Compatibility".

[23] http://www.adobe.com/products/flashplayer/

6.3.2 Simplicity

It was important that Talos be simple enough to be used effectively by people with little or no programming experience. Although the toolkit was aimed mainly for the research community, content writers (e.g., tourist book writers) wishing their content to be "heard" through a multimodal medium could also benefit from it. Consequently, it was designed so that users are not required to learn any language (scripting or otherwise) to write scripts that describe the behaviours of ECAs or complete other tasks with the system. Any programming functionality should be hidden unless it is explicitly called by an advanced user. System processes should be purely visual, like putting the pieces of a Lego game in place. For example, teaching the ECA what questions to expect from the user should be as simple as typing the text in the relevant field. Any other system parameters needed to process the question should be set automatically.

6.3.3 Modularity

Because of substantial user demand, technologies in the mobile space are constantly evolving. For example, only recently computer vision - a component that is currently not available in Talos - has enabled the identification of landmarks (e.g., LookTel[24], 2013) without the use of any barcodes. Therefore, in order to keep Talos relevant to the needs of its stakeholders, its architecture should enable the easy integration of new components, either third party components or custom made. To achieve fast and easy integration and smooth interaction with the core modules, the toolkit APIs should facilitate the development of new modules, and/or integration of third party modules written in a variety of programming languages. Developers should be provided with APIs, along with guidelines for developing compatible system modules.

6.3.4 Expressiveness

Animating ECAs that move and gesticulate appropriately with spoken text is a very difficult task. I believe that these nonverbal behaviours should be realised by enabling

[24] http://www.looktel.com/

the human author to combine predefined gestures from an animation library, and not be automatically created by the toolkit. A developer who has full control over the ECA gestures can realise a more diverse (and believable) series of behaviours than could otherwise be created using automatic animation. In addition, the developer can precisely define where, and for how long, these behaviours would appear in the content, a process that is still imperfect in automatic animation. However, manually crafting predefined animations is a daunting task. Therefore, Talos should provide a tool that streamlines the process, by making it easy to capture the required motions and modify them on demand.

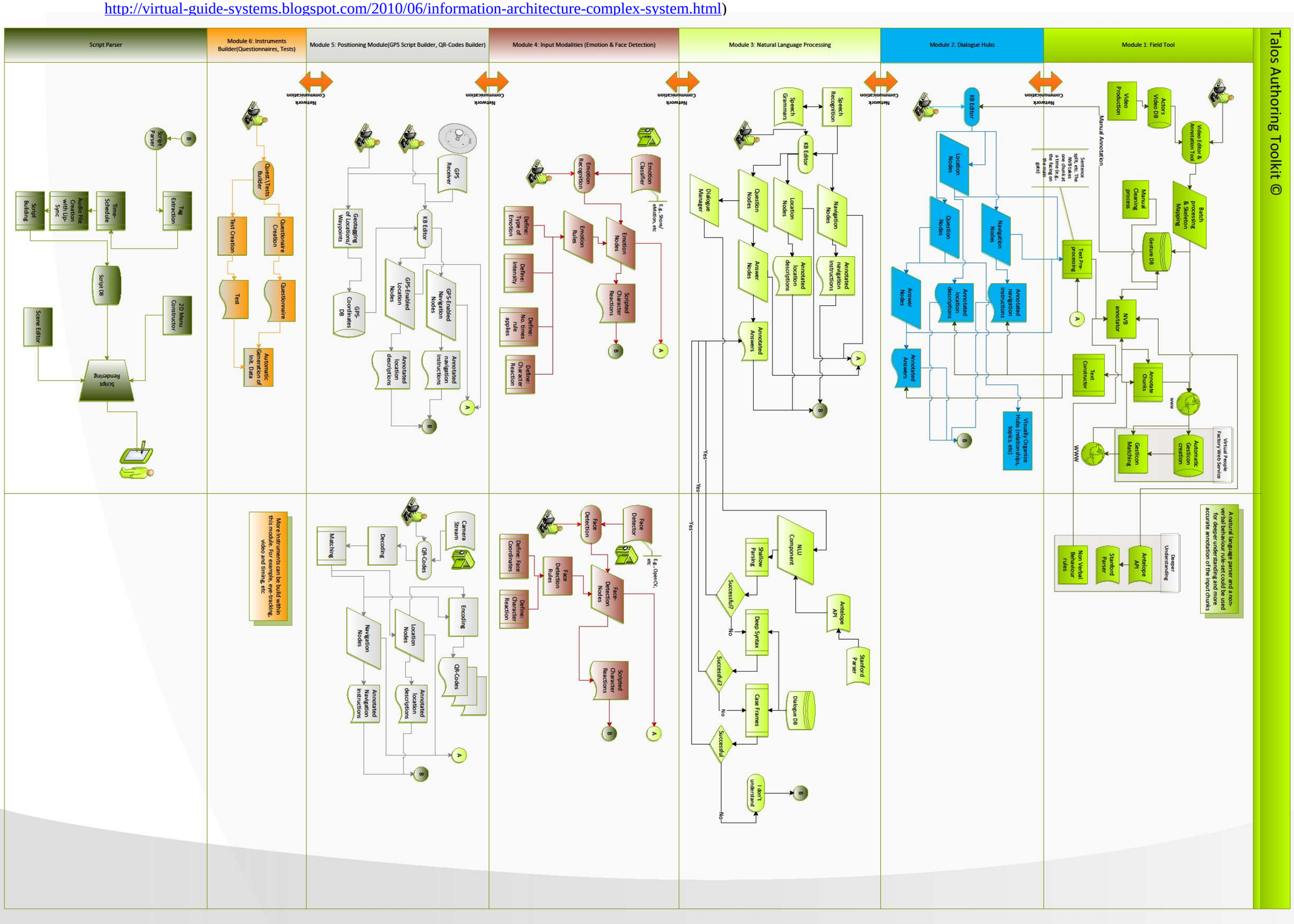
6.3.5 Synchronization

The Talos toolkit should provide two modes of nonverbal – automatic speech synchronization and manual speech synchronization. In automatic mode, the proper behaviours are assigned to the content by the system; to save the human author time and effort of repeated tasks (e.g., assigning behaviours to common words like “Hello” or “Welcome”). Though automatic, this mode still requires the inspection of the content creator to ensure that the system has properly assigned the behaviours. In manual mode, the author is solely responsible for tagging the content with the proper nonverbal behaviours from a gesture library. In both modes, though, the system should automatically create the proper scripts for synchronised verbal and nonverbal ECA behaviour.

6.3.6 Natural Language (NL) Robustness

Because of the variability of natural language, lack of fluency in speech, and potential errors introduced by the automatic speech recognition, Talos’ mapping of users’ questions onto system responses must be very robust. A simple pattern matching system like AIML (Wallace, R. 2003) will not work, because to process structural variations of a single sentence requires as many patterns as there are syntactic alternatives, which leads to very large databases with question-answer pairs that are difficult to handle and modify. As a general principle, if the system has a question that is even slightly similar to the input, the answer assigned to the system’s question should be returned as the system’s output.

Figure 6.3: The architecture of the Talos toolkit
(For a high resolution image please see
http://virtual-guide-systems.blogspot.com/2010/06/information-architecture-complex-system.html)



6.4 Architecture of Talos

The architecture of Talos is shown in Figure 6.3. It consists of six modules, which are discussed in detail below. Some of these ideas were implemented in this research work, but implementing the full UI toolkit is a task better suited for a team of designers and developers.

6.4.1 Module 1: Field Tool

The Talos field tool is a markerless optical motion-capturing and editing system, aimed at streamlining the process of creating the gesture database needed for ECA nonverbal behaviour realization. The tool operates in three phases: a human motion production phase, a motion editing phase, and an automatic skeleton mapping phase. Producing human motion begins by collecting a video corpus from a real human. A human could be a professional tour guide or someone playing the role of a tour guide. In the motion editing phase, the captured videos are edited to extract the desired gesture clips and are fully annotated with location and gesture information (i.e., keywords that seem relevant to the current gesture and context). The videos are then processed and the resulting motion data are automatically mapped onto the skeleton of the ECA. The generated gestures may or may not contain spatial information. In the first case, the tool allows developers to manually adjust the parameters of the motion file to create custom made pointing-to gestures.

The NVP (i.e., Nonverbal Processor) annotator of the tool, once invoked by the "Dialogue Hubs" module, automatically determines the appropriate gestures for an input chunk of text (e.g., a chunk from a location description) using a gesture lexicon (or gesticon). The gesticon is constructed automatically from the user's annotations of motion files and it contains two main types of data: the name of the gesture motion file and a number of keywords that activate it. An alternative option to keyword matching, that requires further exploration, is the deeper analysis of the input chunks for their

syntactic dependencies. A nonverbal behaviour rule set could be used to specify the associations between syntactic dependencies and nonverbal motion files.

Currently the tool handles only arm and body gestures, but the same motion capture approach can be extended to facial expressions. By capturing and classifying motion capture facial expressions, the ECA can display scripted emotions without the need to build complex dynamic emotional models.

6.4.2 Module 2: Dialogue Hubs

The dialogue hubs module facilitates the construction of menu-based dialogues for interacting with ECAs. The module supports the creation of knowledge bases (KBs) consisted of nodes. In order to support information variability, each KB can be constructed to represent a specific variety of information for the same attraction (e.g., "Historical", "Architectural", etc.). Each KB may contain four types of nodes: location nodes, navigation nodes, question nodes and answer nodes. Question and Answer nodes and their relationships can be visually organized into several levels. As discussed previously, nonverbal annotation of location, navigation and answer nodes can be generated either automatically or manually. The module can output multimodal information in real-time upon the developer's request. For example, the content developer may wish to see and modify the ECA's reaction to a certain question. S/he can click on the desired question and watch the ECA execute the annotated text.

6.4.3 Module 3: Natural Language Processing

Similar to dialogue hubs, this module supports the development of knowledge bases consisting of nodes. However, as opposed to menu-based dialogues, this module accepts speech as an input modality and applies deep linguistic processing to a question before it matches it with an answer. In addition, the module facilitates the development of speech recognition grammars, i.e., structured collection of words or phrases that the ECA should recognize. The natural language understanding feature of this module has been partially

realized and is discussed in Section 6.5. Like the "Dialogue Hubs" module, the ECA can execute annotated content in real-time, but it can now be activated using the developer's voice.

6.4.4 Module 4: Input Modalities

This module has two main functions: first, it is a real-time emotion classifier and face detector. Emotion detection is achieved using a third party facial expression analysis system that classifies the face into the set of "prototypical" emotions such as happy, sad, angry, etc. (Ekman *et al.* 1969) (e.g., the Fraunhofer SHORE engine (Fraunhofer Institute, 2010)). Face detection is achieved using a custom-made API for the Intel's open-source face detection algorithm (OpenCV, 2011) that returns coordinates to the module for the position of the face relative to the camera. The emotion classifier and the face detector are provided as alternative inputs and cannot be used in conjunction. Second, the module aids the development of behavioural nodes to specify what input the ECA should recognize and the kind of behaviour it should exhibit as a reaction to the input. For example, the following dialogue illustrates a scripted ECA reaction once the parameters (i.e., an emotion of certain intensity was detected, occurring once during a presentation) of the relevant behavioural rule have been satisfied.

Guide: I can see you are confused!
User: Yes, Thanks for stopping, I am really losing you; your presentation is too technical I would like to hear something simpler.

The face detection feature of the module was fully implemented in one of the prototypes. The idea was to mimic how a human guide would react if a group member would wander away while s/he is giving a presentation about an attraction. The following scenario is extracted from the prototype (see §7.1 for a discussion).

Agent: This gate is the only gate to the area of the castle……..
The user wanders away from the guide without paying attention

Agent (stops the flow of the presentation): I am about to start my presentation would you like to come closer? (Once the guide detects that the user is closer, she resumes the normal flow the presentation)….. As I was about to say, this gate is the only gate to the area of the castle.……

6.4.5 Module 5: Positioning Module

The positioning module facilitates the development of location-sensitive ECA scripts. These are scripts that are triggered based on the current location of the user either automatically or semi-automatically. The module currently supports two positioning technologies: GPS (Global Positioning System) detection and QR (Quick Recognition) code Recognition (see below).

In the first case, the content developer creates knowledge bases with location-sensitive nodes with location-specific content and navigation content. These nodes are annotated with pre-recorded GPS coordinates (i.e., longitude and latitude coordinates) and are triggered automatically when the user is within the range of the coordinates. The coordinate range is determined manually during a pre-processing stage, where the developer geo-tags the desired locations and waypoints. In the GPS mode, both pre-determined and dynamic routes are possible. As long as the system has the relevant coordinates and their associative locations and waypoints, it can build routes based on the user's selection of the desired destinations.

A QR Code is a low-cost, two-dimensional bar code that can store up to 4,000 alphanumeric characters and is readable by simple cameras. With these features, a QR Code can be the ideal solution for navigation and information retrieval in environments where the GPS signals are lost (e.g., structurally-dense physical environments) and other sensor-based position estimation technologies are not possible. The use of QR Codes in information retrieval is demonstrated in one of the prototypes, discussed in experiment two (see Chapter 7). Very briefly, the prototype decodes QR Code tagged locations and triggers the relevant ECA script(s). Geo-tagging the locations was as simple as printing

the QR Code and attaching it in a prominent place. The same process can be applied for geo-tagging any alternative routes between locations. QR Codes strategically placed on a route can be used to trigger ECA navigation instructions that take into account nearby landmarks. In addition, bright-red QR Codes can be used as location beacons for people who may have deviated from the planned route. If the user gets lost, s/he may locate the nearest beacon QR Code and retrieve navigation corrections from it, i.e., multimodal instructions on how to return to the last known landmark of the planned route.

6.4.6 Module 6: Instruments Building

Generating and administrating the proper instruments for scientifically evaluating ECA-based mobile applications can be an impractical and error prone process. This can be particularly true when it comes to field evaluation, where environmental conditions can significantly affect the way people use the instruments. Then, getting the first indications of the success or failure of the evaluation can take hours of data extraction and manual calculations. This module of Talos enables ECA researchers to build the necessary instruments with ease and automatically administrate them after the experimental session. To facilitate instrument building, the module provides templates for various instruments such as interactive questionnaires, multiple-choice or keyword based memory tests, etc. Once the user has completed giving his/her feedback, the module automatically checks the instruments for errors (e.g., any missed questions) and generates the initial numeric data without the researcher's intervention.

6.4.7 Script Parser

The script parser is an internal component of Talos, available to all the relevant modules. The script parser enables the multimodality of the ECA by automatically creating the scripts needed for synchronizing the spoken text with nonverbal animations. Furthermore, the script parser processes custom-made tags that control other elements in the character's environment (e.g., its background). First, the script parser extracts the

animation tags from the content. Following is a snippet of content used in one of the prototypes:

\book =<back,C17> Please \book=<anim,back> head your way back to the main street of the castle. After making \book=<anim,portello6> at the first opportunity two left turns.

Figure 6.4: Example of a Talos multimodal script

It then calculates a time schedule by accessing the text-to-speech (TTS) engine and the total duration of the character's build-in animation sequence. Finally, it enables the development of GUI elements such as buttons, display windows, etc. and it smoothly renders the final output on the screen of the mobile device. A UI version of the script parser was successfully implemented for the purpose of building the scripts needed for the prototypes.

6.4.8 Scripting Language

For advanced developers, Talos offers a scripting language named GSL (Guide Scripting Language) for greater control over the toolkit. This language is generated automatically by the toolkit, but it can be modified by the user on demand. It is an XML-compliant language; much like AIML (Artificial Intelligence Modelling Language) (Wallace, R. 2003) designed to describe the tasks performed with the toolkit UI, as discussed above. This data is encapsulated with a series of system tags, with the option for custom tags. For example, the knowledge-base containing the introduction state of a dialogue generates the following script:

```
<GSL>
<topic>Introduce</topic>
<response>
    <user_input>How can I start the tour</user_input>
    <speech>Simply \book=<anim,next> tap the button next</speech>
    </response>
</GSL>
```

Figure 6.5: Example of a Talos GSL Script

The most useful feature of GSL is that it enables advanced developers to create plugins made accessible by associative tags. Developers can implement an infinite number of new plugins from real time access to web services (e.g., Google search, Google translate, etc.) to custom made solutions for tailored functionality. Plugins and the toolkit API for developing new modules are the two options for extending the functionality of the toolkit to match the requirements of any mobile project with ECAs.

6.5 Talos Language Processing Component: Design and Development

In this section, I focus on the natural language processing component of the Talos toolkit. The component was developed mainly for the evaluation goals of one of the prototypes (see §8.1 of Chapter 8). However, it could also benefit the general virtual human research community by providing a more robust and linguistically-motivated solution than the widely used AIML (Artificial Intelligence Modelling Language). The component uses a four-layered approach to map an input string from the user to an appropriate response in the database. Figure 6.6 shows the workflow of the current implementation of the module. At the first-layer of processing is a third-party system, called the **V**irtual **P**eople **F**actory (Virtual People Factory, 2013). The designer defines the knowledge contained in a domain by entering (in plain English) pairs of questions & answers in the VPF's database. It provides the first processing layer for the input by computing its similarity to a question (called a trigger) in the database. Once the trigger is found, the system responds with the answer (called a speech) for the trigger (Dickerson *et al.* 2005). A major step in the matching process is the use of a list of global keywords. These are the most important words used by triggers globally and are extracted in real-time when the designer enters the question & answer pairs in the database. However, as the keywords are not annotated with part of speech (POS) information, VPF fails to distinguish ambiguities between triggers that contain the same global keywords. For example, consider the following two triggers: a) shall we begin the tour[25]? b) Can we tour now?

[25] The component POS tagger produces the following annotation: Shall/NNP (Proper Noun) we/PRP (Pronoun) begin/VBP (Verb) the/DT (Determiner) tour/NN (Noun).

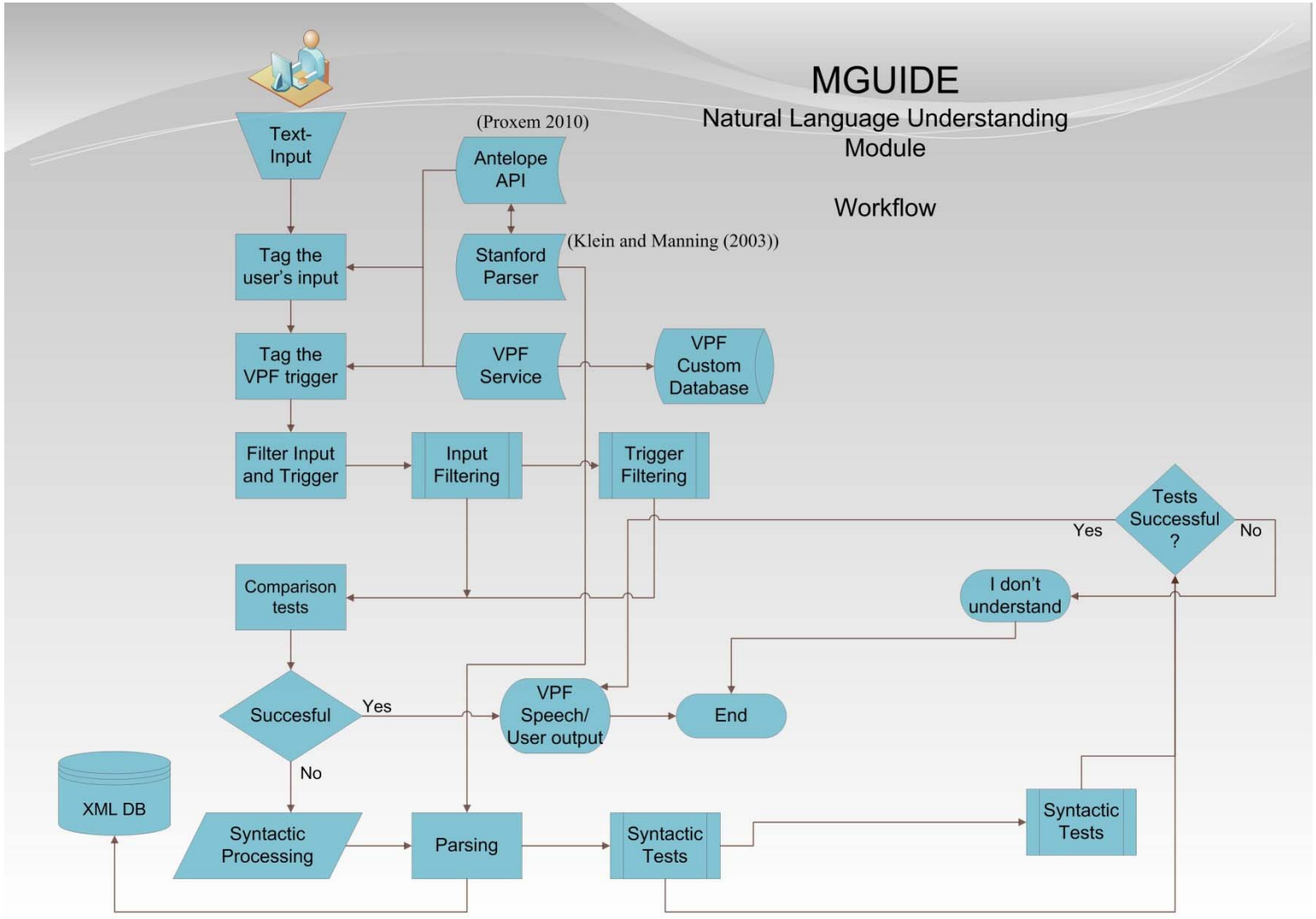


Figure 6.6: The workflow of Talos NLU module (current implementation)

The third layer performs a deep syntactic analysis on the input string. It parses the input for predicates and deep syntax dependencies[26], and searches for the best match of the parsed input against phrases in the database. The matching process involves several comparisons/matching tests with incrementally relaxed conditions (see Appendix B for a code snippet of these tests). This ensures that if at least some of the predicate arguments of the input and the database are the same (or similar), matching will be successful. Once a match is found, the phrase is passed to the VPF[27] for an exact match. If this step fails, the system either does not have a response, or it did not understand the question the way the user asked it. Hence, it replies with a generic off-topic response (e.g., *"I do not understand please rephrase or move to a different topic"*). The evaluation of the algorithm as a whole, along with a comparison performance of the three processing layers is presented in Chapter 8 (see §8.2.2.1 of Chapter 8).

In the full Talos architecture (see Figure 6.3 Component 3: Natural Language Processing), there is a fourth processing layer between the above two layers that was not included in the current module's implementation. It performs shallow semantic analysis of the input text. A shallow form of semantic representation is a case-form analysis, which identifies the sentence's predicate (e.g., a verb) and its thematic roles[28] (e.g., AGENT, EXPERIENCER, etc.). In a few words, this process assigns a "who did what to whom, when, where, why, and how" to the input sentence. Currently, the module's semantic component is not "mature" enough to be used in a real dialogue application. It uses an open-source semantic parser (Proxem 2010) which is highly experimental. However, even if the parser is improved in future versions, it is unlikely that it will become powerful enough to resolve accurately the natural language ambiguities even in limited domains. Consider for instance, the utterance *"I want more information about the church"*. The subject "I" can be considered either as an AGENT (i.e., who performs the action) or the EXPERIENCER (who receives the result of the action) of the predicate, so there are two distinct case-frames. I developed an experimental semantic processing stage in the current algorithm (discussed above)

[26] The deep syntax parsing for the utterance "Shall we begin the tour?" looks something like this: begin (Subject: we DirectObject: tour)
[27] http://www.virtualpeoplefactory.com
[28] Thematic role is the semantic relationship between a predicate (e.g., a verb) and an argument (e.g., the noun phrases) of a sentence (Glottopedia, 2010)

that addresses this problem (within limits). In particular, it uses a predefined library of valid case-frames in the domain of the prototypes (e.g., frame_want, frame_see, etc.) in order to automatically cut any invalid interpretations. With this constraint, it then searches for specific thematic roles (and their values) to help make sense what is being discussed. For example, the previous sentence maps to "frame_want" with thematic roles and values: PATIENT: Information, GOAL: Church. Once the same case name and a case component with the same label and value match, the utterance for that frame is returned. Of course, more research and development is needed to refine this stage, but it can currently match correctly a range of questions (and paraphrases) to the corresponding frames in the sample database. The code for this stage will be released as open source, along with the rest of the algorithm.

A last component of the NLP module of Talos is the dialogue manager. It is modelled as a Hierarchical Tasks Decomposition process: acts[29], topics, subtopics, and their associative trigger templates. Trigger templates are framed-like structures with slots representing a trigger phrase's case frame, predicate argument structures and POS keywords. For each trigger phrase in the database, a separate template is defined.

This process is being carried out semi-automatically, where the system generates the templates automatically and the designer manually corrects any failed or multiple interpretations of the trigger phrase. This hierarchy allows the system to keep the context as it has detailed information on what has been activated at each level of the conversation (e.g. POS keyword(s)). Figure 6.7, shows a generated graph for the domain of the prototypes (i.e., a tour of a medieval castle) along with two trigger templates for the phrase *"I want more information about the central gate"*.

[29] I divide the information space into acts (i.e., largest unit of information), topics (i.e., smaller unit of information) and subtopics (i.e., the smallest unit of information). For example in Figure 6.7 "Location A" is the act that includes several topics and subtopics.

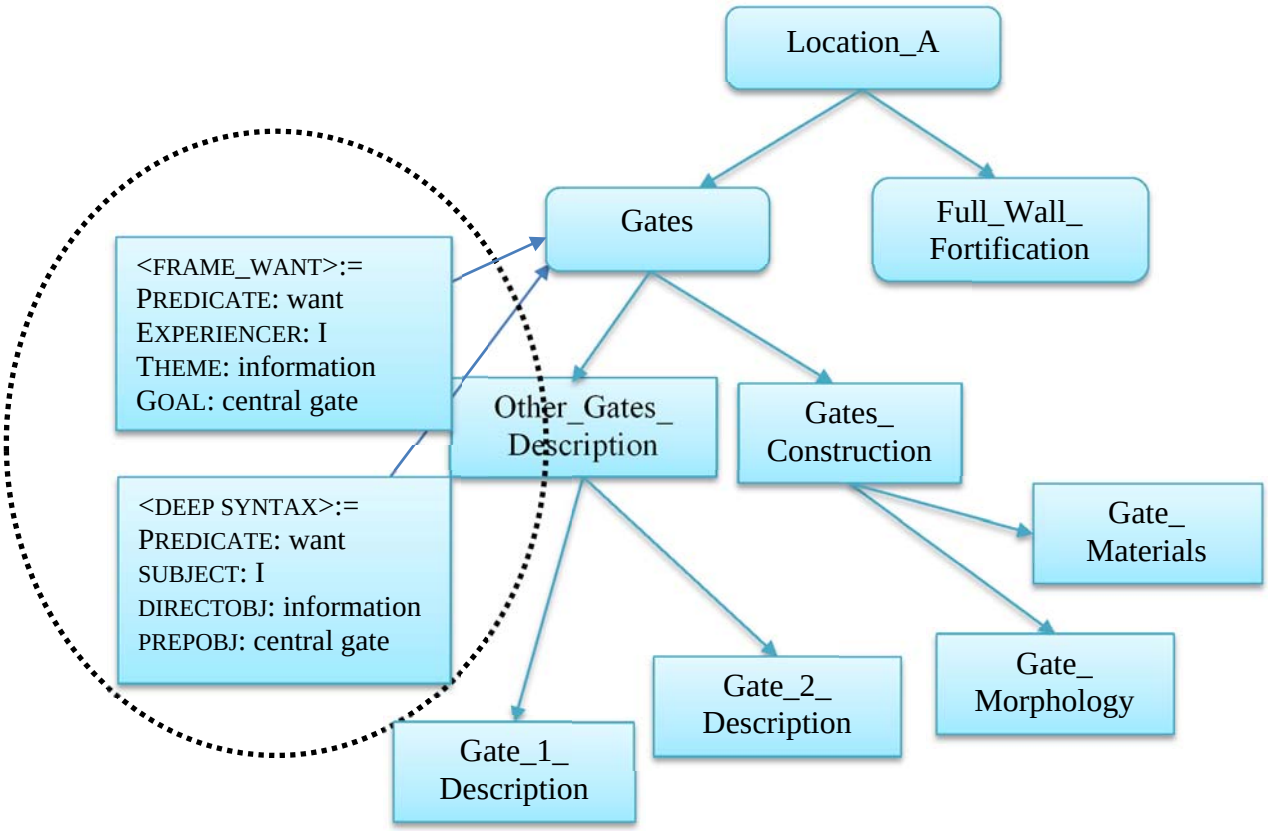


Figure 6.7: A sample graph generated by Talos dialogue manager

Each of the above nodes carries an activation list, where the developer specifies:

1) How many times a node should be activated. For example, the "Introduction" nodes are activated only once. This way, the ECA can understand when the greeting time is over and the real conversation begins.

2) Prioritize the activation of the nodes using a priority value. For example in the graph of Figure 6.7, the node "Gate_Morphology" logically has a greater probability to be next in the discussion than the "Gate_Materials". This is because a user will most likely ask questions about the form of the gate first, before getting into questions about the materials that were used in its construction. This prioritization value for each node is difficult to determine, and should be empirically determined through testing.

3) Define each node activation preconditions. For example, the "Gates_Construction" node can only be activated if the node "Gates" has been activated first.

4) How the system responds when the above conditions are not met. For example if the greeting time is over and the user says hello, a reply could be *"Giannis, how many times are you going to say hello to me"*.

6.5.1 Comparison with other language processing systems

My approach for processing natural language has a weighting keyword matching algorithm (i.e., Virtual People Factory) at its core, but it significantly extends its capabilities. Other language processing systems, available on the web for free are: a) the Personality Forge Engine[30] (Personality Forge, 2013) b) the PandoraBots[31] (PandoraBots, 2013) – the web based implementation of AIML (Wallace, R. 2003). However, there are a number of differences between these systems and VPF.

1) Both "Personality Forge" and "PandoraBots" require rules to be said verbatim to match the input string. On the other hand, VPF uses a matching heuristic to determine the similarity of the input (*"OK, I am ready let's begin the tour"*) to an entry in the script (*"let's begin the tour"*).

2) Because of the above approach in input matching, VPF requires fewer rules to answer the same output in comparison to both "Personality Forge" and "PandoraBots". This has a significant impact on the system's performance and management of scripts for large application domains.

3) VPF, in contrast to "Personality Forge" and "PandoraBots", enables a developer to define "how-well" a rule should match the input and cut-off any matched-rules below that threshold level.

[30] http://www.personalityforge.com/
[31] http://www.pandorabots.com/botmaster/en/home

4) Creating a good script in “PandoraBots” and “Personality Forge” requires extensive knowledge of each engine’s internal scripting language. VPF scripts, on the other hand, are written in plain English.

5) VPF considers the information space in terms of acts (very large chunks of information) and topics (smaller chunks of information). “PandoraBots” divides the information space only into topics. “Personality Forge” does not create subsets of input data as topics of conversation readily.

6) VPF provides an easy-to-use system to deal with the system’s failed responses. The absence of a similar system is perhaps the biggest weakness of “Personality Forge” as it makes correcting failed output from the system a very difficult task.

7) VPF endows developers with full control over a response of the system. However in “Personality Forge”, the AI engine takes control of the system’s output with random responses quite often being produced.

8) The VPF in contrast to “Personality Forge” offers a reliable web environment for development and testing of Virtual Humans. The low data transfer speeds of “Personality Forge” limit the usefulness of that service.

9) Contrary to “Personality Forge”, VPF offers a free and easy-to-use API (Application Programmable Interface) for integration into applications.

10) Although “Personality Forge” uses world list wildcard rules, these are not associated with a single word and therefore are not automatically reusable throughout a script. VPF offers a simple but very intuitive “Synonym-List” finder that automatically associates the chosen keywords with synonym lists for the entire script.

In the desktop space, AIML (Wallace, R. 2003) seems to be the best representation of contemporary language processing. There are several desktop implementations

(e.g., AIMLpad[32], AIMLBot[33] and others), but they differ little from PandoraBots. The NPCEditor of the ICT Virtual Human toolkit (discussed in Appendix A.1), appears to be the closest solution. However, to perform even the simplest of tasks (e.g., mapping whatever is not known to an off topic answer), requires tweaking a number of complex system parameters.

6.5.2 Future Work

Developing and releasing the full NLP module of the Talos toolkit is a long term goal. However, its current implementation i.e., the three-tier NLU algorithm, can be released as open-source immediately for the benefits of the virtual human research community. My current work involves developing an editor and a simple API, to allow developers to integrate with ease, Question & Answering (Q-A) functionality to their applications. The editor allows the designer to map sample questions-to-answers in a simple and straightforward way. It updates its internal databases (see Figure 6.7) automatically, while the user enters the question-answer pairs in the editor. The lexical information (e.g., predicate synonyms) required by the algorithm, are provided by the designer in the settings panel of the editor. Other more advanced features, like the threshold applied in each script, are accessible via the web interface of the VPF system. The API is compatible with all the recent Windows operating systems (OS) integrated development environments (IDEs), which I hope will encourage the wider dissemination of the algorithm in developing systems. I welcome collaborations on the further development of the editor and other modules of the Talos toolkit.

6.6 Conclusion

This chapter has provided an overview of the Talos authoring toolkit, my novel solution for making the development of conversational virtual humans for mobile applications significantly easier. The usability of two similar toolkits was evaluated by an expert (see Appendix A.1) and a set of practical recommendations was generated that should guide any future attempts to develop the toolkit.

[32] The AIMLpad is available for download (free of charge) at http://program-n.sourceforge.net/
[33] The AIMLBot is available for download (free of charge) at http://sourceforge.net/projects/aimlbot/files/

From the full toolkit, I have partially developed the language processing component for the purposes of the evaluation of one of the prototypes. I have compared my approach for language processing with similar publicly available/open source systems. I have plans to release the current three-tier algorithm as open source, along with an easy to use editor and an API for application integration.

In the next chapter, a number of empirical studies with systems generated with the prototypes of Talos are discussed. The empirical evidence collected, together with substantial accumulation of personal knowledge on designing and developing ECA-based mobile systems, was amalgamated into a series of recommendations on how to improve the user experience of ECA interactions in mobile environments.

Chapter 7 ECA Visual Presence Studies

This chapter discusses three studies that aim to provide some insights into the main question of this research, i.e., whether the presence of a multimodal Embodied Conversational Agent (ECA) improves the cognitive accessibility and usability of a mobile tour guide system and enhances the user experience of such systems. The first experiment examined the impact of an ECA on the ability of participants to navigate routes and to retain information from locations in a real-tourist attraction, namely the castle of Monemvasia, Greece. The second experiment examined the problem of information retention separately from that of navigation, while the third vice versa. All experiments were conducted in simulated mobile conditions using high-resolution panoramic images (experiments one and two) and high-resolution video-clips assembled into two interactive applications (experiment three).

For each experiment, I initially give an overview of the experiment and the initial hypotheses. These hypotheses were derived from an exploratory study that took place in the actual castle of Monemvasia. Then an outline of the design of each experiment is provided. Finally, I discuss the analyses that I performed on the collected data and the conclusions that I reached based on the generated evidence.

EXPERIMENT 1

7.1 Overview

This study was designed to evaluate the impact of the presence of a multimodal ECA on the cognitive accessibility and usability of a mobile tourist guide interface, as well as the quality of the user experience. The experiment was conducted in a lab under simulated mobile conditions, with an experimenter present whose task was to observe the session. The agent's visual presence (present vs. absent) and order of systems (ECA-present then ECA-absent vs. vice versa) were manipulated to observe any practice effects. To represent the routes and attractions the user would visit in

the real castle of Monemvasia, I used high-quality panoramic photographs[34] with resolution 6000 x 3000 pixels (see Figure 7.1) which I assembled into two interactive applications. Each tour was projected on the wall through a projector attached to a Sony Vaio FZ21Z[35] laptop with resolution 1280 x 800 pixels. Users interacted with the application through a wireless mouse attached to the laptop. To move from location to location, users had to click on-screen buttons. Although the setting was designed to be simple enough to use without any previous training, I allowed each user maximum two minutes[36] at the beginning of each experiment to become familiar with the use of the panorama. As opposed to other studies where the mere presence of an ECA was manipulated (e.g., Miksatko *et al.* 2010), I sought to compare a number of modalities for outputting information from the tour guide system.

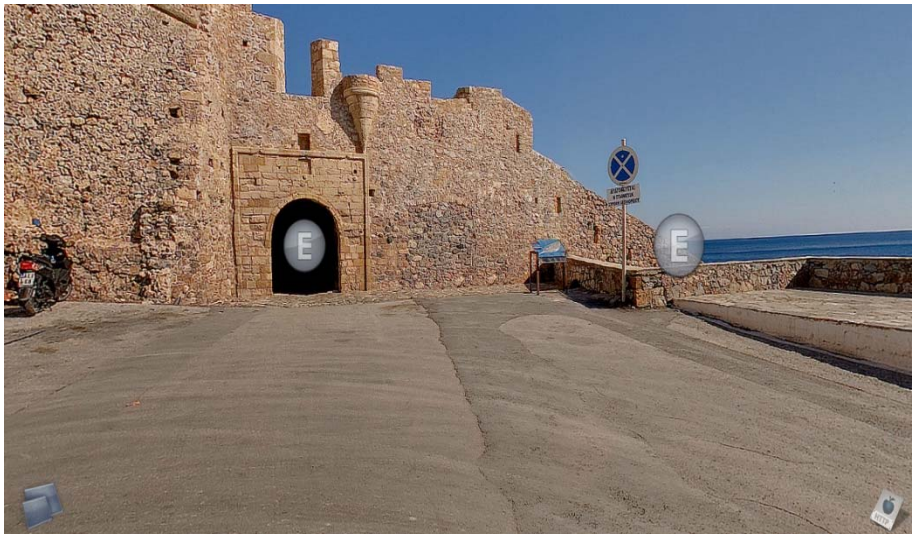


Figure 7.1: A screenshot of one of the two interactive panoramic applications (Users could explore the 3D scene and click on E (Next) to continue)

In particular, I compared (see Figure 7.2) an ECA that uses a range of full-body gestures and facial expressions, and its anthropomorphic presence, with a non-ECA system that uses voice and text-subtitles. The argument for this approach is that

[34] A video of one of the prototypes can be found at http://www.youtube.com/watch?v=hh_02KoG8M4&feature=related

[35] The laptop's full technical specifications can be found at http://www.laptopsdirect.co.uk/Sony_VAIO_FZ21Z_VGN-FZ21Z/version.asp

[36] All participants confirmed that they felt comfortable with the use of panoramas within the allocated time

even if the mere presence of an ECA has no significant effects on motivation and learning performance (Miksatko *et al.* 2010) then the additional communication modalities may, at least, enhance performance, as the information presentation would be enhanced by being transmitted through multiple channels (see *H4: "The information retention enhancement hypothesis"*).

Figure 7.2: The system with the ECA (left side) and the system without the ECA (right side)

Based on the findings of an exploratory study (see Appendix C), I generated the following hypotheses to test in this study. The hypotheses reflect the possible (though contradictory) outcomes of this experiment and an explanation of the most likely reasons for them.

Information Retention:

H3: The information retention degradation hypothesis: The presence of a multimodal ECA decreases retention performance in the user by: a) distracting users away from the information presentation on each of the locations and b) stimulating users to explore (through question-asking) the information available about a location in more-depth. The more content the user explores through question-answering the more time s/he spends to complete a tour, the more content would be generated and the less s/he will most likely remember the content presented about a location.

H4: The information retention enhancement hypothesis: The presence of a multimodal ECA increases retention performance in the user, for instance, because

given a system capable of personalizing the information presentation on each of the locations (and given a visual agent capable of generating appropriate non-verbal behaviours to accompany the linguistic information and guide the user's attention focus), it renders the interaction with the system more smoothly, enhances the quality of the user's information processing, thus potentially supporting greater retention.

Navigation Performance:

H5: The navigation enhancement hypothesis: The presence of a multimodal ECA enhances the user's ability to navigate, for instance by helping him/her in understanding the underlying structure of the physical space and, hence, allow him/her to better navigate himself (see the second tenet of distributed cognition in §3.2.3 of Chapter 3, for the supporting theory).

H6: The navigation degradation hypothesis: The presence of a multimodal ECA degrades the user's ability to navigate, for instance by failing to correctly convey landmark information (e.g., because the iconic gestures are out of synchronization with the speech), thus distracting users from locating the landmark in the physical environment.

7.2 Experimental Design

This section presents the design of experiment one. First, I provide an overview of the participants (age, gender, etc.) and the software/hardware equipment they used (including a detailed overview of the ECA features). Then, I present the tasks participants were assigned to complete, and the conditions under which they completed them.

7.2.1 Participants

In total, twenty-one able-bodied participants (both males and females), took part in this study. Three of the participants were used in a pilot study, to ensure that the main experiment would run problem-free. Those three participants completed

exactly the same tasks as the others, but spent overall more time in the lab to discuss improvements and identify any bugs the systems might have. Based on their feedback, I made a number of last-minute improvements in the design of the system's interface and the instruments of research. The remaining eighteen participants (see Table 7.1 and D.1.1 in Appendix D for the full participants' details) were randomly assigned to the experimental conditions. Although every effort was made to ensure that the prototypes were bug-free, some issues remained. On some occasions participants experienced problems with the synchronization of content i.e., the 3D board (see §7.2.3) was slightly out of synchronization with the spoken content. These fluctuations were noted by the experimenter and corrected only for the participants that were affected after a session with the systems. In order to avoid over-familiarity with the area of the Monemvasia castle, no participant was either a local resident or had visited the site before. All participants were native Greek speakers, and had a variety of academic and mobile-device backgrounds.

Order of systems	Participants	Age (Mean)	Std. Deviation	Gender (M/F)
ECA Present vs. ECA Absent	9 participants	36.8	7.7	4/5
ECA Absent vs. ECA Present	9 participants	32.6	5.2	5/4

Table 7.1: Table of participants in experiment one

7.2.2 Software and Equipment

The design requirements generated by the exploratory study (see Appendix C), were effectively translated into two high-fidelity prototypes. Additional requirements were gathered through observation of real humans giving tours of archaeological attractions. From this "requirements backlog", I only implemented an ECA attention mechanism that I thought could significantly impact the user's retention of information from presentations about the locations of the castle.

The ECA-based system features a simple interface with photographs of landmarks/locations users would encounter in the tour as the system background, and the following user interface objects (see Figure 7.2a): a dialogue-window; a control-window; and a face tracking window (invisible to the user). The dialogue window displays a menu of text-phrases/questions for the user to select from, dynamically updated based on the user's selection and the current context. The text-phrases cover a broad range of possible questions/clarifications a user could ask after a presentation about an attraction. The non-ECA system (see Figure 7.2b) features the control and dialogue windows and a subtitle window that displays the contents of the system. A text-to-speech (TTS) engine "reads" the content, while highlighting each word of the text.

7.2.3 ECA

The same ECA model was used in all six experiments of this research work. The ECA has a photo-realistic 3D appearance (donated by Haptek Inc.[37]) capable of eye-blinking and movement of its mouth in synchronization with the synthesized voice. In addition it features:

- More than 2000 handcrafted gestures: It uses deictic gestures to indicate a point in space, for example, to show the user a specific point on the background photograph of a location/landmark (or to simply show the participants which way to go) and other gestures to emphases specific parts of the information provided.

- The ability to utilize additional multimedia information (on a 3D board) to enhance further the information provided about an attraction.

- Automated awareness of the user: The ECA can dynamically evoke the attention of the user back to the tour and the content is presenting at the particular instance. This was achieved through the implementation of a computer vision module that allowed the ECA to detect the position of the user's face and react accordingly (e.g., interrupt a presentation and request

[37] The Haptek corporate homepage at: http://www.haptek.com

the user's attention back) if it could not see the user, or if the user holds the device too far away from his/her face.

7.2.4 Task

Participants were administered individually using the Samsung Q1 Ultra-Premium tablet[38] device with resolution 1024 x 600 pixels. After a brief explanation (identical for all participants) about the purpose of the experiment, participants began the task, which was to navigate one of two pre-selected routes in turn (see Figure 7.3 and Figure 7.4 for route segments), visiting a number of locations and uncovering information of interest about particular attractions. Participants were randomly assigned to the routes. The routes were distinct, but similar in terms of experimenter-estimated navigation difficulty. Participants were asked to perform this task, once using an interface with the ECA, and the other using an interface without the ECA.

Figure 7.3: Images from a segment of the first route

Figure 7.4: Images from a segment of the second route

At the beginning of the experimental session, participants had to provide their personal information (name, gender and age), personalize the system's content to

[38] The tablet's full specifications can be found at http://www.samsung.com/us/pdf/UMPC_LR.pdf

their preferences by selecting one of the available information scenarios, and parameterize various features of the agent and the system (e.g., the ECA appearance, volume, etc.). The information scenarios (i.e., "Architecture", "History" and "Biographical" details) were based on the cultural and historical background of the castle. Participants could freely choose the scenario they wanted. I offered this option rather than forcing a specific scenario, because I wanted to evaluate the impact of personalized content (see *H4: "The information retention enhancement hypothesis"*) instead of forcing of participants to experience content they might not have been interested in. The content in all scenarios was of equal length and complexity. However, while every care was taken to simplify the content using terminology specific to the architecture domain could not have been avoided.

After that, the ECA (which supposedly has knowledge about the area) appeared either as a virtual character (left side of Figure 7.2) or a disembodied voice with a subtitle window (depending on the testing condition) (right side of Figure 7.2), and proactively provided information about the locations users as visited using the panoramic applications.

In order to visit a location, users had to make the correct navigation decisions on a number of junctions first. Each system provided a photograph of the junction's major landmark and a brief verbal description about which way to go (based on the landmark) and what to do next. For example, the speech instruction for the first photograph in Figure 7.3 was, *"Please walk towards the main gate of the castle. Passing through it you will find yourself in a dark catacomb. Continue your way in the catacomb and make the first turn on your right."* To respond to this instruction, the user had to click on the correct button on the panoramic scene that loads the next segment of the route. Once the next segment was loaded, the user had to tap the button next to the control window to load a photograph of a new landmark and an audio instruction on what to do next.

After visiting a number of landmarks, the user arrived at the location, where s/he could ask the system to start its presentation using the dialogue window, or to take some time to have a look around the attraction first. The remainder of the photographs in Figure 7.3 illustrates the landmarks that the user had to follow in

order to get to the first location of the tour. After the presentation of information, the user could ask the ECA questions from a dialogue window with a menu of dynamic text phrases. When the presentation (and any possible questions) was completed, the user was able to move to the next location by tapping "next". After participants had visited all the locations, they had to take a retention test, complete three questionnaires on their experiences with the systems and finally participate in a short interview.

7.2.5 Conditions

The condition where participants experienced the system with the fully multimodal ECA (see left side of Figure 7.2) was defined as ECA-present. The condition where the participants experienced the system with the voice and the subtitles (and without the ECA) (see right side of Figure 7.2) was defined as ECA-absent.

Participants (n = 18)	**ECA Present**	**ECA Absent**
1 – 9 Participants	**Route One** Performance/Questions/ Quest. A + Quest. B + Quest. C/Observation Notes/Interview Responses	**Route Two** Performance/Questions/Quest. A + Quest. B + Quest. C/ Observation Notes Interview Responses
10 – 18 Participants	**Route Two** Performance/Questions/Quest. A + Quest. B + Quest. C/ Observation Notes Interview Responses	**Route One** Performance/Questions/Quest. A + Quest. B + Quest. C/ Observation Notes Interview Responses

Table 7.2: Experimental design of experiment one

In the first order (abbreviated as P/A) participants experienced the system with the ECA on the first route, and then the system without the ECA on the second route. In the second order (abbreviated as A/P) participants experienced the system without the ECA on the first route, and then the system with the ECA on the second route. The ECA's visual presence (absent and present) and order of systems (ECA-present

then ECA-absent vs. vice versa) were manipulated as independent variables. The dependent variables were performance (i.e., scores on the retention test, time taken to complete a route, and frequency of getting lost). In addition the following data were recorded, the number of questions asked per location for a tour, the users' ratings on the assessment questionnaires, the observations of users' behaviour made by the experimenter, and the responses to the questions posed in the short interview. The variable type of agent was manipulated within-subjects (see Table 7.2), whereas the variable order of systems between-subjects. The eighteen participants were assigned at random half in the first order, and the other half in the second order.

7.3 Measures and Methods

Both objective and subjective measures were used. Towards the more objective end of the scale, I measured the following:

- Time: The total time to complete a tour (in seconds) was recorded using automated device logs.

- Quantity of question-asking per participant: The total number of questions participants asked was recorded using automated device logs.

- Frequency of getting lost: A participant was defined as lost if, in a panoramic scene that represented a junction, s/he would click the wrong button on the panoramic (and hence would deviate from the planned route) and the experimenter had to intervene in order to get him/her back onto the route. The navigational errors were noted by the experimenter using a paper notebook.

- Retention of the presented information: The participants' retention performance was measured using electronic retention tests. The test was administered on the same laptop that was used to display the panoramas. It used a fill-in-the-blanks approach (see Table D.1.11 in Appendix D for the retention test) that required the participant to fill-in a number of missing

words in sentences, carefully selected from the presentations about each location. For each sentence, participants also had to rate the confidence of their answer on a ten-point scale (1 = completely at random, 5 = not so confident, 10 = totally confident).

- The observation notes on users' behaviour made by the experimenter.

The primary subjective variables in this experiment were:

- The responses to the individual items of the three electronic questionnaires: The three questionnaires were administered on the laptop used to display the panoramic castle tours.

  The first questionnaire, assessed the cognitive accessibility of the systems and used a mixed yes/no and seven-point Likert scale (1=strongly disagree, 7 = strongly agree) items. It addressed the effectiveness of the ECA in guiding the user's attention focus to specific objects in each location and a number of aspects of the user's satisfaction (see §7.4 of this Chapter for a summary of the questionnaire).

  The second questionnaire addressed a number of usability dimensions of the systems (see §7.4 of this Chapter for a summary of the questionnaire).

  The third questionnaire examined a number of qualities of the agents (see §7.4 of this chapter for a summary of the questionnaire). The experimenter made a number of behavioural observations in a paper notebook while users interacted with the prototypes.

- The answers to the post-task interview questions posed by the experimenter: The questions posed in the interview were open-ended and provided participants an opportunity to give their impressions about the systems, the ECA and offer suggestions, about what should be improved in future versions. Each interview lasted 5-10 minutes.

7.4 Results and Discussion

Having presented the design of experiment one, this section focuses on the discussion of its results. First, I discuss the results of the objective measures. These are the data collected for: "time to complete a tour", "questions asked", "navigation errors" and "retention performance". Then, I discuss the results of the subjective measurements. These are the data collected with the four questionnaires: "Object Recognition Questionnaire", "Cognitive Accessibility Questionnaire", "Usability Questionnaire", "ECA-Specific Questionnaire" and the participants' feedback from the post-task interviews.

Performance Measures

Table 7.3, shows the mean results for the objective user task performance (see Tables D.1.2, D.1.3, D.1.4 and D.1.5 in Appendix D for more details). First, a 2 x 2 ANOVA test taking time as the dependent variable and type of ECA and order of systems as the independent variables, showed that the average time (in seconds), to do each task did not differ as a function of either main effect, namely type of ECA or order of systems. There was a significant interaction between type of ECA and order of systems ($F(1, 32) = 31.588$; $p < .001$) (see Figure 7.5). No other ANOVA comparisons reached significance level.

Participants (n = 18)	**ECA Absent**	**Std. Deviation (Absent)**	**ECA Present**	**Std. Deviation (Present)**
Time to complete a tour	1390.1	339.1	1431.4	388.0
Questions Asked	9.6	8.5	8.7	5.2
Navigation Errors	1.6	1.0	1.5	1.6
Retention performance	29.6	20.21	27.4	18.5

Table 7.3: Objective user task performance

The interaction of type of ECA and order of systems was analysed using simple main effects analysis. The variation of order of systems influenced the time

performance of participants using the system with the ECA ($F(1, 32) = 25.416$; $p < .001$), and the system without the ECA ($F(1, 32) = 8.450$; $p < .01$). A closer look of the descriptive statistics (see Table 7.4) reveals that participants using the system with the ECA spent overall less time in the second order (mean A/P = 1120.3), than in the first (mean P/A = 1742.5). On the other hand, participants using the system without the ECA, spent more time in the second order (mean A/P = 1569.5), than in the first order (mean P/A = 1210.7). This is clearly a practice effect. As the design of the systems was similar and the type of ECA did not have any effect, participants were already familiar with the systems the second time they completed a tour with either of the two systems (with or without the ECA).

ECA	P/A (n = 9)	Std. Deviation	A/P (n = 9)	Std. Deviation
Present	1742.55	293.49	1120.3	126.4
Absent	1210.7	338.5	1569.5	239.6

Table 7.4: Time as a function of ECA and order of systems

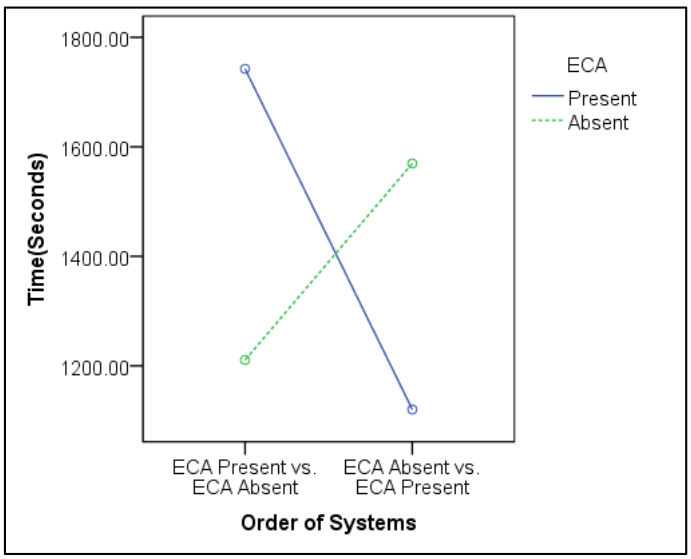


Figure 7.5: The interaction of time for ECA and order of systems

Second, in order to explore the issue of question-answering further, I carefully inspected the device logs. A 2 x 2 ANOVA taking the "number of questions asked" as a dependent variable and order of systems and type of ECA as independent showed no significant effects for either the type of ECA or order of systems. No

significant interactions were found either. A correlation between the questions asked and the participants' retention performances returned no statistically significant results. The experimenter noted that most participants skipped the part of Q&A with the systems. As they later explained in the interviews, if the presentations were shorter to match their personal time constraint, they would have asked the ECA more questions. Therefore, it is evident that the time constraint of visitors is an important factor towards general user acceptance (and possibly comprehension) of cultural heritage content. For example, short-stay visitors would most likely appreciate short presentations and to be given the ability to ask any questions later if they wish. Long-stay visitors, on the other hand, would most likely be uninterested in short term presentations as they would like to experience in full what the castle has to offer.

With respect to the number of times that participants got lost, a 2 x 2 ANOVA taking the navigational errors as the dependent variable and order of systems and type of ECA as independent, did not show any significant main effects of the two variables. However there was a significant interaction between the type of ECA and the order of systems ($F(1, 32) = 10.240$; $p < .01$) (see Figure 7.6). This interaction was further analysed using a simple main effects analysis. The variation of order of systems significantly influenced the navigation performance of participants using the system without the ECA ($F(1, 32) = 6.969$; $p < .05$) but not the system with the ECA. The navigation errors of participants using the system with the ECA did not differ much across the two order conditions (mean P/A = 1.00 vs. mean A/P = 2.1) (see Table 7.5). However, the participants using the system without the ECA took significantly less correct navigation decisions in the first order than in the second order (mean P/A = 2.4 vs. mean A/P = 0.88). This may be because the participants had to follow different routes across the two order conditions.

ECA	P/A (n = 9)	Std. Deviation	A/P (n = 9)	Std. Deviation
Present	1.0	1.11	2.1	1.96
Absent	2.4	0.88	0.88	0.60

Table 7.5: Navigation errors as a function of ECA and order of systems

The two routes were of similar navigation difficulty but could not, of course, be made exactly equal in difficulty. Therefore, a more effective experimental design would have been to mix up the group of participants so each group tried both routes with each of the systems (with and without the ECA). However, this was hard to achieve in practice as I did not have access to a pool of participants that would be available on demand and for as long as I needed them to evaluate the systems on both routes.

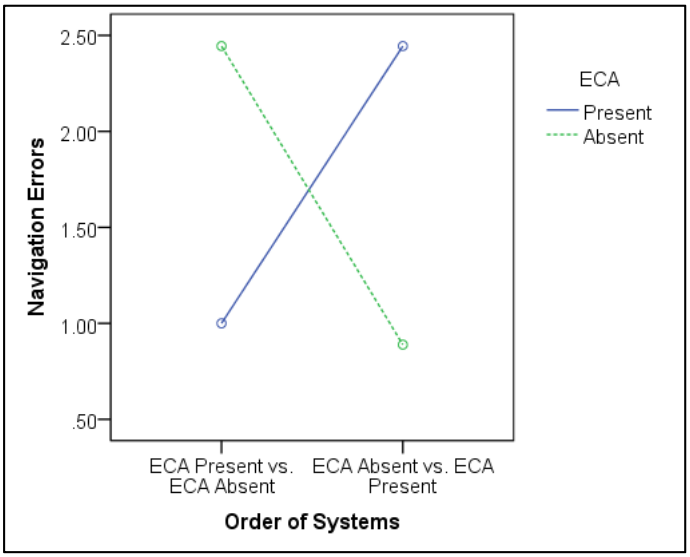


Figure 7.6: The interaction of navigation errors for ECA and order of systems

Participants using the system without the ECA performed better on the route that required slightly easier navigation decisions. On the other hand, participants using the system with the ECA were not considerably affected by the choice of the route. A reasonable explanation is that the deictic gestures used by the ECA helped participants in disambiguating the navigation instructions given by the systems. This fits with the navigation enhancement hypothesis (see H5) that a multimodal ECA can enhance the user's ability to navigate. However, this enhancement effect is related to the complexity of the route as participants navigated the route that required them to take less complex decisions (with either the system with the ECA or the system without the ECA) with no errors (mean ECA-present = 2.1 vs. mean ECA-absent = 2.4).

Finally, a series of 2 x 2 ANOVAs taking retention scores and confidence as the dependent variables, and type of scenario (i.e., "Architecture", "History", and "Biographical"), order of systems and type of ECA as independent variables did not produce any significant main effects. This suggests that the presence of a multimodal ECA has no effect on retention performance which invalidates both my information retention hypotheses (see H3 and H4). However, there was a significant interaction between the type of ECA and order of systems ($F(1, 32) = 7.165$; $p < .05$) (see Figure 7.7). This interaction was further analysed using simple main effect analysis. It revealed that the variation of order of systems significantly influenced the retention performance of participants using the system without the ECA ($F(1, 32) = 4.639$; $p < .05$) but not the participants using the system with the ECA (see Table 7.6).

ECA	P/A (n = 9)	Std. Deviation	A/P (n = 9)	Std. Deviation
Present	34.3	19.0	20.4	19.8
Absent	20.5	16.2	38.8	16.7

Table 7.6: Retention score as a function of ECA and order of systems

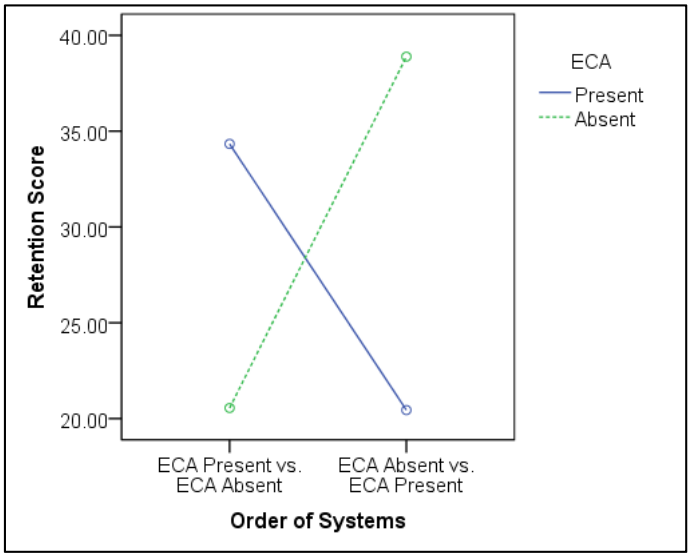


Figure 7.7: The interaction of retention score for ECA and order of systems

This evidence provides a strong indication that my information retention enhancement hypothesis (see H4) could be valid. This is because the modalities

used by the system with the ECA were more consistent in presenting different content about the locations of the castle than the modalities used by the system without the ECA. Given this presentation consistency, plus the observation that the overall difference between the participants' retention performance was minimal (see Table 7.3), if the modalities used by the system with the ECA were improved, it is highly possible that the ECA could actually benefit retention performance instead of degrading it.

Subjective Assessment

Object Recognition Questionnaire

The table below (see Table 7.7) shows the total number of the physical objects participants recognized and the objects that they missed during the presentations across the two orders (order P/A vs. order A/P). I conducted a chi-squared test which returned no significant results within or between the conditions.

Order of systems	ECA Present (Yes/No)	ECA Absent (Yes/No)
ECA Present/ ECA Absent	54 / 27	61 / 20
ECA Absent/ ECA Present	60 / 21	53 / 28
Total (Y/N)	114/48	114/48

Table 7.7: The physical object (Y/N) recognition results

Furthermore, there was no difference between the total number of the objects participants successfully recognised and the objects they missed, using either the system with the ECA or the system without the ECA. Hence, it can be said that both systems were effective in guiding the participants' attention focus to physical objects of a location.

Cognitive Accessibility Questionnaire

The Likert cognitive accessibility questionnaire consists of nine questions-sets (see Table D.1.7 in Appendix D for the full questionnaire), each asking questions reflecting the requirements of each zone the Simplex II model (see §3.2.2 of

Chapter 3). The first set includes six items, and it was designed to assess the complexity, learnability, consistency and self-organization requirements of the navigation and information tasks. The second set consists of four items. It was designed to assess the participants' perception of the prototypes' output modalities (visual, auditory and textual) in terms of satisfaction and understanding (e.g., understanding the system dialogues, visibility of the device screen etc.). The third set examined how the participants perceived the feedback (visual, auditory and textual) given by the prototypes in terms of sensory, timing, relevance and memory requirements. This question set consists of six items (e.g., synchronization of information, relevance of information to the castle, etc.). The fourth set assessed the demands the prototypes placed on the participant's working memory, and it consists of four items (e.g., the system should respond properly when the participant is confused/overloaded with information, the amount of information to hold in mind when using the system, etc.). The fifth set evaluated the emotional impact of the prototypes, and it consists of four items (e.g., disappointing, annoying, etc.). The sixth set assessed the impact of the prototypes on the participants' long-term memory in terms of the learnability of the task, and in relation to their existing knowledge. This question set includes four items (e.g., relation of information to the participant's personal interests, ease of learning of the information and routes, etc.). The seventh set assessed how effective participants could access the underlying structure of the information and navigation tasks, and it includes four items (e.g., ease of building a mental "picture" of the route and the presented information, etc.). The eighth set evaluated how the participants perceived their responses to the prototype's requests in terms of rationality and response support. It consists of four items (e.g., allowances for navigation errors, frequency of wrongly retained information and/or navigation decisions, etc.). The ninth and final set evaluated how the participants felt about their responses to the prototypes' requests and the support they received in order to respond properly. It includes four items (e.g., frequency of getting support to proceed from location to location, support needed to learn the information presented about a location, etc.).

A cronbach alpha on each of the nine sets showed the following (see Table 7.8). Most alpha results, range from "questionable" up to "good", except the alpha for the question set five that is below "unacceptable". A review of the "Item-Total"

statistics showed that if Item 24 (*"The system is fun to use"*) is removed, the alpha increases significantly (a=0.666). This question will be reviewed or completely removed from future versions of the questionnaire.

Q.Set	Alpha
1	0.658
2	0.796
3	0.530
4	0.546
5	0.139
6	0.632
7	0.691
8	0.635
9	0.730

Table 7.8: Cronbach alphas of the cognitive accessibility questionnaire

A series of 2 x 2 ANOVAs taking each questionnaire item as a dependent variable and order of systems, type of ECA and scenario as independent showed a significant effect of order of systems on the following five questionnaire items (see Table 7.9):

- Item 12 (*"The information provided (i.e., speech, gestures, facial expressions, and images) by the system is not correctly synchronised"*) ($F(1, 32) = 19.174$; $p < .001$)
- Item 13 (*"The information provided (i.e., speech, gestures, facial expressions, and images) by the system is not clear")* ($F(1, 32) = 4.699$; $p < .05$)
- Item 23 (*"The design of the system is not serious enough"*) ($F(1, 32) = 4.472$; $p < .05$)
- Item 30 (*"I find it hard to understand the structure of the content presented by the systems"*) ($F(1, 32) = 4.997$; $p < .05$)
- Item 39 (*"The system does not give me any support to learn the information about a location"*) ($F(1, 32) = 6.197$; $p < .05$)

Cognitive Accessibility		**ECA**			
Item	**Order of systems**	**Absent**	**Present**	**AVG**	**Std. Deviation**
12	P/A	1.2	1.1	1.1	0.3
	A/P	2.0	2.6	2.3	1.0
13	P/A	1.4	1.7	1.6	0.9
	A/P	2.0	3.1	2.5	1.6
23	P/A	1.1	1.2	1.1	0.3
	A/P	1.8	2.2	2.0	1.6
30	P/A	2.3	1.6	2.0	1.2
	A/P	2.4	3.6	3.0	1.6
39	P/A	1.3	1.3	1.3	0.4
	A/P	2.2	2.5	2.3	1.6

Table 7.9: Cognitive accessibility questionnaire items with significant order of systems effects

Finally, I found a highly significant effect of scenario on several questionnaire items:

- Item 2 (*"It's difficult to learn the tasks of navigation and information extraction"*) ($F(2, 31) = 3.597$; $p < .05$)
- Item 7 (*"I cannot clearly see the screen of the system"*) ($F(2, 31) = 3.749$; $p < .05$)
- Item 8 (*"I cannot hear the dialogues of the system")* ($F(2, 31) = 27.880$; $p < .001$)
- Item 10 (*"It's difficult to understand the dialogues used by the system")* $F(2, 31) = 13.127$; $p < .001$)
- Item 13 (*"The information provided (i.e., speech, gestures, facial expressions, and images) by the system is not clear")* ($F(2, 31) = 12.193$; $p < .001$)
- Item 25 (*"The design of the system makes it difficult to learn what I have to learn to use it correctly"* ($F(2, 31) = 4.568$; $p < .05$)

- Item 28 *("The information scenarios should be related better with my personal interests"*) (F(2, 31) = 4.766; p < .05)
- Item 38 (*"I always have to ask for Giannis help to navigate from location to location"*) (F(2, 33) = 4.180; p = < .05)
- Item 40 *("I never know the correct navigation instructions to get to my destination")* (F(2, 31) = 10.925; p < .001)

A review of the descriptive statistics (see Table 7.10) for Item 12 (*"The information provided (i.e., speech, gestures, facial expressions, and images) by the system is not correctly synchronised"*) shows that participants in the first order perceived the content presented by the two systems as better synchronised (mean P/A = 1.1), than the participants in the second order (mean A/P = 2.3). The minor corrections I made to the fluctuations in synchronicity of all scenario scripts do not provide an adequate explanation for this effect. If there had been an influence of the corrections in synchronicity participants in the second order would have rated the systems better for the particular item or the same as in the first order. On the contrary, they rated the systems worse. The most likely explanation for this order effect is a group difference in the way participants perceived the ECA-based systems. The second group included participants with game testing experience (see D1.1 in Appendix D for the list of participants) that may have demanded more in terms of content synchronicity than what the systems (with and without the ECA) could actually offer.

Order of systems	Scenario	Mean	Std. Deviation
ECA Present vs. ECA Absent	Architecture	1.5	0.7
	History	1.1	0.3
	Total	1.1	0.3
ECA Absent vs. ECA Present	Architecture	3.2	0.5
	History	2.2	1.0
	Biographical	1.0	0.0
	Total	2.3	1.0

Table 7.10: Item 12 mean ratings as a function of order of systems and scenario

The same pattern is repeated in the effect of order of systems on Item 13 (*"The information provided (i.e., speech, gestures, facial expressions, and images) by the system is not clear")*. If participants in the second order demanded more in terms of content synchronicity, then it is to be expected that they thought the content is less clear (mean A/P = 2.5), than the participants in the first order (mean P/A = 1.6).

Then, for the effect of order of systems on Item 23 (*"The design of the system is not serious enough"*) (see Table 7.9) participants in the second order perceived the design of the systems as less serious (mean A/P = 2.0) than the participants in the first order (mean P/A = 1.1). Participants, who selected the "Architecture" scenario in the second order, rated this item higher than the participants who selected all the other scenarios in both orders. The most likely explanation is that the participants thought that the design of the systems is not serious enough to handle the complexity of the "Architecture" scenario. This pattern for architecture seems to be repeated in Item 30 (*"I find it hard to understand the structure of the content presented by the systems"*) and Item 39 (*"The system does not give me any support to learn the information about a location"*) and in all of the significant effects for the scenario.

With regards to the significant effect of order of systems on Item 30 (*"I find it hard to understand the structure of the content presented by the systems"*) Table 7.9 shows that participants in the second order had more difficulty in following the structure of the content (mean A/P = 3.0), than the participants in the first order (mean P/A = 2.0). The last significant effect of order of systems on Item 39 (*"The system does not give me any support to learn the information about a location"*) follows the same pattern, with participants in the second order perceiving the systems as less helpful (mean A/P = 2.3), than the participants in the first order (mean P/A = 1.30). The explanation for both order effects can be found in the scenarios participants selected in the second order. Participants who selected the "Architecture" scenario rated both questionnaire items higher than the participants who selected all the other scenarios in both orders. It is evident that those participants struggled to handle the complexity of information provided by the systems in the "Architecture" scenario.

The above pattern for "Architecture" is also repeated in the significant effects for scenario on the questionnaire items 2, 7, 8, 10, 13, 25, 28, 38 and 40. A series of post-hoc comparisons using the Tukey HSD (Hsu, 1996) test indicated that the mean scores for the "Architecture" scenario were significantly different from the scores for the other three scenarios. Participants who selected the "Architecture" scenario rated all of the questionnaire items higher than the participants who selected the other three scenarios.

Usability Questionnaire

Apart from the effects of ECA (and order of systems) on aspects of the user's perceived cognitive accessibility of the systems, I asked participants on a Likert survey on whether they found the systems usable enough. For more details on the questionnaire, see Table D.1.8 in Appendix D. As before, I measured the reliability of the aspects that each of the four questions sets was designed to measure. The first set contains six items, and it was designed to assess the ease of use (e.g., difficulty of use, understanding of terminology, etc.) of the prototypes. The second set assessed how efficiently users were able to accomplish the tasks (e.g., usefulness, attentiveness, efficiency, etc.), and it includes six items. The third set measured the likability of the systems (e.g., degree of engagement, innovation, etc.), and it consists of six items. The fourth and final set contains eight items, and it was designed to evaluate the participants' feelings (e.g., confusing, frustrating, etc.) towards the prototypes. The results showed that the four question sets were overall "poor", with the first item being "unacceptable" (see Table 7.11).

To identify the cause of the generally low alphas, I reviewed the "Item-Total" statistics of each question-set for problematic questions. I found the following problematic questions that will be reviewed or completely removed in future versions of the usability questionnaire: Item 4 (*"The system uses terms understandable and familiar to me")* in group 1 (a=0.248); Item 15 (*"I thought that my conversation with the system was unnatural"*) in group 3 (a = 0.794); and Item 24 (*"Refreshing"*) in group 4 (a = 0.635).

Q.Set	Alpha
1	-0.334
2	0.036
3	0.432
4	0.440

Table 7.11: Cronbach alphas of the usability questionnaire

A 2 x 2 ANOVA was carried out to determine whether the variables I manipulated had any effect on the ratings. It turns out that there was a significant difference between the conditions for specific questionnaire items. In particular, there was a significant effect of scenario on some questionnaire items (Items 1, 2, 8, 12, 19, 21, 24, 25, and 26), that was because of the issues participants experienced with the "Architecture" scenario.

I also found a significant effect of order of systems on the following questionnaire items (see Table 7.12):

- Item 5 *("The system has too many choices")* ($F(1, 32) = 5.394$; $p < .05$)
- Item 24 (*"Refreshing"*) ($F(1, 32) = 6.649$; $p < .05$).

Usability		**ECA**			
Item	**Order of systems**	**Absent**	**Present**	**AVG**	**Std. Deviation**
5	P/A	4.7	4.4	4.6	1.6
	A/P	2.8	4.0	3.4	1.3
24	P/A	5.7	5.6	5.7	1.2
	A/P	4.3	4.3	4.3	1.8

Table 7.12: Usability questionnaire items with significant order of systems effects

With regard to Item 5 *("The system has too many choices")*, Table 7.12 shows that participants in the first order thought that the systems have more choices (mean P/A = 4.6), than the participants in the second order (mean A/P = 3.4). The

descriptive statistics (see Table 7.12) reveal that the participants rated the item similarly across the two order conditions with the system with the ECA (mean P/A = 4.4 vs. mean A/P = 4.0), but not with the system without the ECA (mean P/A = 4.7 vs. mean A/P = 2.8). Most likely participants thought that the system with the ECA constitutes an additional choice and the slightly different route across the two orders, did not affect their perception. On the other hand, as participants with the system without the ECA had to follow a slightly easier route in the second order, their ratings were affected accordingly. The significant effect for Item 24 (*"Refreshing"*) shows that participants in the first order perceived the systems as more refreshing (mean P/A = 5.7) than the participants in the second order (mean A/P = 4.3). The cause of this effect is clearly the issues participants experienced with the "Architecture" scenario. As more participants in the second order chose this scenario than in the first, the low ratings for this item are to be expected.

ECA-Specific Questionnaire

Finally, in the agent-specific questionnaire I asked a range of questions covering the following four aspects of the two agents (see Table D.1.9 in Appendix D for the full questionnaire). The first set includes eight items and it was designed to evaluate various aspects of the agent's behaviour (e.g., intelligence, competence, etc.). The second set assessed various aspects of the agent's voice (e.g., clarity, appropriateness, etc.) and includes four items. The third set includes seven items and assessed various aspects of the agent's appearance (e.g., realism, lip-synchronization, etc.). The fourth set assessed various aspects of the agent's effectiveness in assisting participants to complete the task (e.g., disambiguation of information, erroneous situations, etc.) and includes four items. The fifth and final set evaluated the body and face features of the agent (e.g., realism, synchronization, etc.) and includes eight items. The appearance and body and facial gestures group of questions applied only to the visual system and therefore participants did not answer the questions in the condition without the ECA.

As before, I measured the alpha of these questions sets (see Table 7.13). Apart from the question sets three and five, the reliability results for all other sets are "unacceptable". This indicates a problem with specific questions in each group.

Q.Set	Alpha
1	0.216
2	-0.318
3	0.655
4	-0.052
5	0.627

Table 7.13: Cronbach alphas of the ECA questionnaire

Revising the "Item-Total" statistics for all questionnaires-sets, I found the following problematic questions: Item 6 (*"The virtual guide was emotionless"*) in group 1 (a=0.309); Item 9 (*"I liked the voice of the virtual guide"*) in group 2 (a = 0.544); Item 14 (*"I would prefer a more realistic virtual guide"*) in group 3 (a = 0.723); Item 21 (*"The virtual guide should help me in erroneous situations (e.g., when I am lost in a route)"*) in group 4 (a = 0.050); and finally Item 28 (*"I liked the guide's body language"*) in group 5 (a = 0.696). No other problematic questions were found. These questions will be revisited or completely removed in future versions of the questionnaire.

A series of 2 x 2 ANOVAs taking each questionnaire item as a dependent variable, and order of systems, type of ECA, and scenario as independent showed a significant effect of order on questionnaire Item 3 (*"I thought that the virtual guide was intelligent"*) ($F(1, 32) = 5.311$; $p < .05$) (see Table 7.14). No significant interactions between the variables were found. It also showed a significant effect of scenario on the following items:

- Item 3 (*"I thought that the virtual guide was intelligent"*) ($F(2, 31) = 9.182$; $p < .01$)
- Item 9 (*"The methods of information presentation (i.e., voice, images, gestures and face expressions) are many and confuse me. I would like a simpler system (e.g., with voice or text)"*) ($F(2, 31) = 26.429$; $p < .001$) and
- Item 11 (*"The voice of the virtual guide is not suitable for this system"*) ($F(2, 31) = 20.759$; $p < .001$)

ECA-Specific		ECA			
Item	**Order of systems**	**Absent**	**Present**	**AVG**	**Std. Deviation**
3	P/A	1.5	1.4	1.5	0.8
	A/P	2.1	2.8	2.5	1.6

Table 7.14: ECA questionnaire items with significant order of systems effects

The descriptive statistics for the significant order of systems effect on Item 3 (*"I thought that the virtual guide was intelligent"*) (see Table 7.14), show that participants in the second order thought that the virtual guide was more intelligent (mean A/P = 2.5) than the participants in the first order (mean P/A = 1.5). As discussed below, the most likely explanation of this effect is the complexity of the "Architecture" scenario. This most likely made participants think that the virtual guide presenting it is more intelligent than the virtual guide in the other two information scenarios. As more participants in the second order chose the "Architecture" scenario than in the first, the higher ratings for the avatar's intelligence are to be expected.

A post-hoc comparison on questionnaire Item 3 (*"I thought that the virtual guide was intelligent"*) using the Tukey's HSD test, showed that between the three information scenarios, there is a significant difference between the "Architecture", "History" and "Biographical" scenarios ($p < .05$). Participants who experienced the "Architecture" content thought both systems more intelligent (mean "Architecture" = 3.8), than those who experience the "History" content (mean "History" = 1.6) and "Biographical" content (mean "Biographical" = 1.5). An additional post-hoc Tukey's test on questionnaire item 11 (*"The voice of the virtual guide is not suitable for this system"*), showed that there is a significant difference between all three information scenarios ($p < .05$). Again, participants who experienced the "Architecture" scenario, had stronger opinions (mean "Architecture" = 4.5) that the voice of the guide is not suitable for the systems than the participants in all the other scenarios (mean "History" = 1.6, mean "Biographical" = 1.0). Given the unique nature and complexity of the "Architecture" scenario, these results were to be expected. Then, the descriptive statistics for Item 13 ("*I like the appearance of the*

virtual guide") showed that participants who experienced the "Biographical" scenario liked the appearance of the virtual guide more (mean "Biographical" = 6.0) than the participants in all the other three scenarios (mean "History" = 5.7, mean "Architecture = 5.6). As there was only one participant, who chose this scenario in both orders, it is not safe to make any assumptions about the effects of the "Biographical" scenario on the perception of the guide's appearance.

Finally, participants rated the ability of the character to react to the user's attentiveness of the presentations (Item 19 (*"A virtual guide capable of face-detection and generation of appropriate responses is the minimum interactive feature such a system should have"*)), above average in both orders (mean P/A = 4.4 vs. mean A/P = 4.2). This shows that although the face-detection module was generally slow to respond, participants consider this feature an important element of their overall experience. The experimenter observed that this feature was generally effective in getting the participants' attention back to the presentations, but further work is needed to improve the speed of the ECA responses.

Interview Feedback

During the interview sessions with the participants, I asked whether they had any general comments about the systems. I additionally asked participants, how they thought I could improve the design of the systems, and if they had any comments about the ECA. I asked for suggestions on how to improve the ECA and finally if they had any comments about the questionnaires (e.g., if they would like to expand on any underlying issues found in the questionnaires). To analyse the data I used a custom-made approach (see §5.2.1.2 of Chapter 5), where I looked for corroboration of patterns and/or comments between a feedback and confirmation group. Below, I discuss my findings:

ECA Design:

1) Most of the participants indicated that the presence of the guide is not necessary, as it attracts attention away from the information presentation.

Even improvements on the scale of the movie "Avatar"[39] would make no difference in the retention of the presented information. Those participants preferred the system with the subtitles. (Corroborated pattern by 4 out of 9 participants)

2) Most of the participants indicated that certain features of the ECA could be improved to make it more effective in presenting information. Features include: (Corroborated pattern by 4 out of 9 participants)
   - Better realism (WOW factor in graphics could impact significantly the user).
   - Better lip synchronization.
   - More natural and human-like body and facial language.
   - Slower speech (Text-to-speech is too fast).
   - Replace TTS with a real human voice. This will solve the pronunciation problems of the existing TTS voice.
   - More cultural oriented appearance.
   - The ability to interrupt the avatar while she speaks.

The corroborated patterns above reveal that participants had conflicting views about the utility of the ECA as a guide. Some of the participants thought, that its presence was completely unnecessary and that not even improvements on the scale of the movie "Avatar" would make any difference. On the other hand, some of the participants thought otherwise. They suggested that an ECA with realistic verbal and non-verbal behaviours, and appearance that resembles a typical Greek female guide would be more effective in its task as a guide.

Panorama Design:

1) Participants indicated a number of usability problems with the panoramic applications. For example: (Corroborated pattern by 3 out of 9 participants and also from the participants' observation)
   - Panoramas shared the same type of button for proceeding next and to enter a building (see Figure 7.1). There should be a different type of button for proceeding to the next panoramic than for entering a building.

[39] http://www.imdb.com/title/tt0499549/

- When the user drags the panoramic display its pointer should change to a drag pointer.
- The panoramic display should have a different speed when the user is in the middle of the panorama than when s/he is at the edge.
- There should be an arrow on the panoramic display that points to the item which the system is providing information about.
- Buttons should not be so close to each other

2) The majority of the users thought it was difficult to synchronize their movements between the panoramic applications and the systems (Corroborated pattern by 7 out of 9 participants and also from participants' observations)

The two corroborated patterns above show that participants experienced some issues using the panoramic applications and made a number of suggestions to address them. I believe that the most important one was the inability of the participants to synchronise their movements between the panoramic applications and the systems running on the UMPC device. However, no participant suggested a way to improve the interaction between the panorama and the mobile device. A possible solution would be for participants to interact with the panoramic display using more natural methods than a wireless mouse. For example the Microsoft Kinect allows control of computer applications using hand gestures.

Multimodal Content Design:

1) Existing Content

- The content is highly dense with information and with too many dates (Corroborated pattern by 4 out of 9 participants)

The corroborated pattern above show that participants found the historic content to include too many dates that were difficult to remember and the density of the provided information (regardless of the information scenario) high. Lowering the density of the information, for example by making the content simpler, and using

historic dates only when necessary (e.g., to indicate important historical facts) will address these problems.

- Some of the navigation instructions were not clear enough. (Uncorroborated pattern)

Some participants expressed the views that the navigation instructions were not clear enough. As the pattern was not corroborated, I can safely assume that the navigation instructions were clear enough to understand by the majority of the participants.

2) **New content features:**

- The systems should include content about arbitrary places in the castle. The need for new content features was also corroborated by users in group two. A number of features were suggested, such as additional information about shops, reconstructed houses, local tradition (e.g., dance) (Corroborated comment by 2 out of 9 participants)

Some participants suggested that the system should provide cultural information of general interest (e.g., about local dances, shops, etc.), in addition to the information scenarios provided. This type of information could be an alternative type of information scenario (e.g., information of general interest about the castle).

Mobile Applications Design:

1) **Improvements in the existing design**:

- The design of the dialogue menu is not adequate (see the floating dialogue window (with the blue title bar) on the right picture of Figure 7.2). For example, there are no bullet points to distinguish the questions. The dialogue menu appears like a continuous text and not as selection of questions. It would be easier for users to ask questions using speech recognition. Finally, a question mark button could be used to control the visibility of the dialogue window. (Corroborated pattern by 6 out of 9 participants and also from the participants' observations)

Participants suggested improvements in the design of the dialogue window. In particular, they said that instead of the dialogue window being visible at all times, a button in the form of a question-mark should activate/deactivate it. Furthermore, participants suggested the use of bullet points to distinguish the questions (as they appear as a continuous text). All recommendations can be implemented in future versions of the systems.

- Users should not be given alternatives to processed to the next location. This should be done with either the text or the buttons, and not both. (Uncorroborated comment)

The alternative options to proceed to the next location were criticized by a participant. He commented that the system should provide only one option to proceed to the next location. However as the comment was not corroborated, I can safely assume that it does not reflect the general view of the participants.

2) New features:

- The subtitle system should allow users to go back to the text and read it without the speech. This way the information will become more accessible (in case the user forgets any information). This comment was corroborated by a participant of the second group, who suggested the use of a small rewind button to allow the user to rewind a presentation only for a few seconds. (Corroborated comment by 1 out of 9 participants)

Participants commented on the usability of the subtitle system. They suggested that users should be enabled to go back to the text and read it without having to listen to the speech. Although the comment was corroborated, implementing it would raise other problems. The small size of the subtitle window does not make it ideal for trying to read long content manually (e.g., using a scroll down bar).

- The user should be given an option for a guided or free tour of the castle. On the guided tour, the system works as it is designed at the moment. In the free tour, the user would walk around freely. In this mode, the panorama has

buttons at certain areas indicating that there is information for the particular point. For example, at the main square where the cannon is the system could have a floating menu on the panoramic indicating that there is content (Architectural, Historical or other) for the particular point. Once the user selects what s/he wants, the guide would be synchronized and narrate the particular content. In the real castle of Monemvasia, the guide would have the panorama as a background to allow the user to scroll and discover the points for which the character knows about. (Corroborated comment by 1 out of 9 participants)

- Selective zooming and/or an arrow could effectively replace the ECA's pointing gestures. (Corroborated pattern by 3 out of 9 participants)
- A map should be constantly available to users, so they can better navigate by themselves. (Uncorroborated comment)
- In the real environment of the castle, the panoramas would be dynamically adaptable based on the physical location of the user (angle, zoom etc.) (Uncorroborated comment)

The above feedback, suggest several features that future versions of the systems could have. The first corroborated comment dictates the use of content-enabled objects that would trigger the system to present information. Although the idea is very interesting, the participants gave me contradictory views on where the objects should be placed (e.g., on the panoramas or the background of the character) and whether their content should replace the main narration about the location or simply repeat it. The corroborated pattern and the uncorroborated comments show features that are difficult to implement (e.g., a dynamically adaptable panoramic background) as it would require radical redesigning the systems.

The above data hardly shows any evidence of the influence of the corrections in synchronicity that was made to the system with the ECA and the system without the ECA. In fact, the analysis of the questionnaire results (see Tables 7.9 and 7.10) shows that there was no observable influence of these minor corrections. If this was the case, it was expected that participants in the second group would have rated the systems better or the same as in group one. On the contrary, they rated the systems

worse. Overall, as it is hard to be conclusive about the influence of the synchronicity corrections the results of this experiment should be treated cautiously in the light of subsequent experiments.

Finally, although the above data could be different in a real outdoor environment, it provides a strong indicator of what users would most likely experience in the real castle of Monemvasia. Although the quality of this experience would be affected by the external environmental conditions (e.g., the sun, noise, etc.) and kinaesthetic factors (e.g., having to walk around in the castle) an improved ECA (in terms of presentation modalities) would most likely benefit the user's retention and navigation performance. Then, the influence of external environmental conditions could be minimised with the latest generation of mobile hardware. For example, the current UMPC devices offer very bright screens (visible even under direct sunlight) and high definition audio.

EXPERIMENT 2

In experiment one; I aimed to rigorously keep the information content of the systems consistent across all conditions. However, the impact of an ECA on the user's comprehension of information (positive or negative) could be correlated to the degree of information complexity. For example, the greater the complexity of the content the stronger (positive or negative) the potential impact of the ECA on the user's ability to retain information, and vice versa. This hypothesis suggests a follow-on experiment to include the issue of the information complexity. In particular, I was interested to evaluate the impact of an ECA on the information content as modified by level of information content with two levels of difficulty – technical and simple. I had the following hypotheses about the potential effects:

H7: The degraded retention of complex information hypothesis: The presence of a multimodal ECA has a negative impact on the retention of the technical information, for example, because it adds an extra burden to the already overloaded cognitive resources of the user (because of the complex nature of the technical information), but neither a positive nor a negative effective for the simple information.

H8: The enhanced retention of complex information hypothesis: The presence of a multimodal ECA increases the participant's retention performance with the technical information, for instance because it reduces the cognitive loads (e.g., by rendering the interaction smoother) needed for retaining such information, but has no effect (neither positive nor negative) on the simple information.

7.5 Overview

This experiment was similar to the first, but it was designed to evaluate the impact of the presence of a multimodal ECA on the accessibility of a mobile tour guide system, providing cultural content of variable difficulty. As before, this experiment was conducted in the lab, under simulated mobile conditions. It manipulated the ECA's visual presence (present vs. absent), the difficulty of the content (simple vs. technical content) and order of presentation (simple then technical vs. vice versa).

Figure 7.8: A screenshot of the interactive panoramic application

The panoramic applications used in experiment one were modified to suit the needs of this experiment. In particular, this application included only a limited number of locations and there were no on-screen buttons for users to interact with. An on-screen menu (see Figure 7.8), allowed users to visit the locations in any order they liked. To start the information presentation the system had to decode the name

of the location from a QR-Code (shown as a camera icon in Figure 7.8) embedded in each location the user visited. A QR-Code is a simple two-dimensional bar-code that can be used as a cheap solution for physical location/object identification. In the current implementation the user has to photograph a QR-Code using the integrated camera of the tablet device.

7.6 Design

In this section the design of experiment two is reported. In particular, I first give an overview of the participants and the software tools they used. Following this, I present the task participants completed and the conditions under which they completed the task.

7.6.1 Participants

In total, fourteen users (both males and females) from a variety of age groups participated in this study. The same group of three participants used in experiment one, took part in a short pilot study to ensure that the formal evaluation would run problem-free. The systems and instruments of research were improved based on their feedback. The participants were randomly assigned to two groups of seven (see Table 7.15 and D.2.1 in Appendix D for the full participant details).

Order of presentation	Participants	Age (Mean)	Std. Deviation	Gender (M/F)
Simple vs. Technical	7 Participants	34	2.8	4/3
Technical vs. Simple	7 Participants	33.8	2.5	4/3

Table 7.15: Table of participants in experiment two

None of the participants was either a local-resident or had visited the area before. This was done to avoid over-familiarity with the area. All participants were native Greek speakers and had a variety of academic and mobile-computer backgrounds.

7.6.2 Software and Equipment

For this experiment, I used a stripped-down version of the systems used in experiment one. All navigation-related functionality was removed, and both systems provided information about a limited number of locations in the castle. Each system, used photographs of each of the locations the user would encounter as a background. The ECA could refer to objects in its background and to additional information that appeared in a floating 3D window. A dialogue window provided users with the ability to have a short "get-to-know-each-other" dialogue with the system (e.g., about how to use the system). A control window provided access to the device's on-board camera and to an interactive map of the castle that showed the locations users had to visit. The non-ECA system (see right side of Figure 7.9) features the same interface elements, but instead of an ECA a subtitle window "reads" the system contents while highlighting each word of the text.

Figure 7.9: The system with the ECA (left side) and the system without the ECA (right side)

7.6.3 Task

Participants were told that the goal of this experiment was to investigate different information presentation systems capable of providing content of variable difficulty about attractions in the castle, with respect to their effects on their ability to effectively retain information. They were asked to use an interactive panoramic application to visit four locations in the castle (in any order they like) and retrieve information with varying degree of difficulty (simple then technical or vice versa), once using system A (i.e., with the ECA) and once using the system B (i.e., without

the ECA). The technical content was a technical description of the locations, while the simple content was taken from the information leaflet the castle provides for free to all visitors. The total duration of each tour was not more than 20 minutes. Furthermore, participants were informed that an experimenter would be present in the lab to observe their behaviour while using the system and to provide help if necessary (e.g., if they could not use the camera to photograph a QR-Code). In addition, they were told that after visiting all locations, they would be asked to indicate in a test what they retained from the presentations.

At the beginning of each task, the system asked participants to provide their personal details (i.e., name, gender and age) and to parameterize various features of the agent and the system (e.g., the ECA's appearance, volume, etc.). After that, a computer agent appeared either in the form of an ECA (left side of Figure 7.9) or a disembodied voice with a subtitle window (right side of Figure 7.9). In order to start a presentation, users had to click on a button embedded in each of the locations they visited using the panoramic applications. The button activated a QR-Code that users had to photograph using the device's camera. Once the QR-Code was decoded, the system would present the relevant information about the particular location (simple or technical). After completely uncovering information for all four locations, participants were asked to indicate, on a five-point scale, whether they found the presentations difficult. Next, a retention test was administered which asked questions about the information they heard in each of the locations. Finally, participants were asked to indicate, their perceived cognitive workloads associated with the presentations they experienced with each of the systems, on a seven-point scale questionnaire.

7.6.4 Conditions

The within-subject variables were the type of content (simple vs. technical content) and type of agent (with levels of ECA-absent and ECA-present), while the between-subjects variable was the order of presentation (simple then technical content vs. vice versa) (see Table 7.16). Participants were randomly assigned to the four experimental conditions: 1) ECA present with the simple content vs. ECA

absent with the technical content or 2) ECA present with the technical content vs. ECA absent with the simple content.

Participants (n = 14)	**ECA Present**	**ECA Absent**
1 – 7 Participants	**Simple Content** Retention performance/Difficulty Rating/Subjective Questionnaire	**Technical Content** Retention Performance/Difficulty Rating/Subjective Questionnaire
8 – 14 Participants	**Technical Content** Retention performance/Difficulty Rating/Subjective Questionnaire	**Simple Content** Retention performance/Difficulty Rating/Subjective Questionnaire

Table 7.16: Experimental design of experiment two

7.7 Measures and Methods

The only objective variable in the experiment was the answers to the retention test. The subjective measures were the responses to the items of the questionnaire, and the ratings of the difficulty of the presentations. The retention test used the same fill-in-the-blanks approach as in the previous experiment. The questionnaire items used the same seven-point agree-disagree Likert (1932) format (1=strongly disagree, 7=strongly agree), and measured the perceived cognitive workload, as an indication of how the participants felt about using the systems to uncover information about the specified locations. The questionnaire was based on the Simplex Two theory (discussed in §3.2.2 of Chapter 3) and included items such as, information complexity, ability to remember the information presented, etc.

7.8 Results and Discussion

The following section discusses the results of experiment two. First, I discuss the results of the retention performance measure and difficulty ratings and then, the results of the qualitative measures. The qualitative measures are the data collected through the workload questionnaire, and comments participants made after the completion of the task.

Performance Measures

As mentioned before, I measured the amount information participants recalled from each type of content as an indicator of the effectiveness of each system (ECA - present or ECA-absent) in eliciting recall performance. A series of 2 x 2 ANOVAs, taking the score and confidence as dependent variables, and type of ECA (ECA-present vs. ECA-absent), order of presentation (simple then technical vs. vice versa) and type of content (simple vs. technical) as independent variables did not show any significant effects of any of the independent variables. There were no significant interactions either.

Order of presentation	ECA (Content) (n = 14)	Mean	Std. Deviation
Simple / Technical	Present(Simple)	25.1	13.4
	Absent(Technical)	21.8	10.99
Technical / Simple	Present(Technical)	21.1	11.9
	Absent(Simple)	36.8	22.4

Table 7.17: Mean retention performances

However, (see Table 7.17) the participants' performances were more consistent with content of varying difficulty with the system with the ECA, than with the system without the ECA. Participants using the system with the ECA performed almost the same between the two content conditions (mean S/T = 25.1 vs. mean T/S = 21.1). Those participants that used the system without the ECA performed better with the simple content (mean Simple = 36.8) than with the technical content (mean Technical = 21.8). This is a strong indication that the modalities used by the ECA (voice, gestures, etc.) were more effective in enhancing the participants' ability to retain information of variable difficulty about the locations of the castle than the modalities used in the system without the ECA. This finding invalidates my original hypotheses (see *H7: "The degraded retention of complex information"* and *H8: "The enhanced retention of complex information"*) as the ECA does not result in enhanced or degraded retention performances. In fact, it has no measurable impact on either the simple or the technical information content. Conversely, the variation of the content affected the participants using the system without the ECA. Their

performance was better with the simple content than with the technical content. Hence, I can safely say that the presence of an ECA does not enhance information retention, but it can provide a more consistent method of presentation for cultural content of variable difficulty, than a system without such an artefact on the interface.

Subjective Assessment

Workload Questionnaire

The workload questionnaire consists of nine questions-sets (see Table D.2.4 in appendix D for more details), each asking questions reflecting the requirements of a zone of the Simplex II model (see §3.2.2 of Chapter 3). The first set of items assessed the complexity, learnability, consistency and self-organization requirements of the information task, and it includes six questions. The second set includes four items, and it was designed to assess how participants perceived the output modalities (visual, auditory and textual) of the prototypes in terms of sensory, satisfaction, and understanding (e.g. visibility of the screen, confusion caused by the multiple modalities, etc.). The third set evaluated how the participants perceived the feedback they received from the prototypes in terms of sensory, timing, relevance and memory requirements (e.g., relevance of the output to the environment of the castle, support to photograph correctly the QR-Codes, etc.). This question set includes six items. The fourth question set evaluated the working memory requirements of the prototypes. It includes four items (e.g., amount of information to hold in mind when using the prototypes, how the system should respond when a participant is confused\overloaded with information, etc.). The fifth set assessed the emotional impact of the prototypes (e.g., frustrating, annoying, etc.), and it includes four items. The sixth set assessed how the prototypes impact the participants' long-term memory in terms of the task learnability and in relation to their existing knowledge (e.g., ease of learning of the information, relation of information to the participant's interests, etc.). This question set consists of four items. The seventh set evaluated how effectively participants could access the underlying structure of the information task, and it consists of four items (e.g., simplicity of the presented information, how the structure of the information is presented, etc.). The eighth set assessed how the participants perceived the

rationality of their responses, and how supported they felt during their responses. This question set consists of four items (e.g., allowances for response errors, frequency of response errors i.e., wrongly retained information, etc.). The final set evaluated how the participants perceived their output responses and how supported they felt in order to respond appropriately. It includes four items (e.g., ease of finding the selected locations, support to learn the information provided, etc.).

Q.Set	Alpha
1	0.850
2	0.646
3	0.653
4	-0.029
5	-4.410
6	0.503
7	0.700
8	0.446
9	0.536

Table 7.18: Cronbach alphas of the workload questionnaire

I applied a cronbach alpha on each of these sets, with the goal to identify any questions that did not measure what they were supposed to measure. It is clear from Table 7.18, that the results for question set four and question set five are below "unacceptable" (i.e., alpha is negative). A review of the "Item-Total" statistics of each group showed that the following questionnaire items should be reviewed or completely removed from future versions of the workload questionnaire: Item 18 (*"I find it difficult to remember that I have to photograph a QR-Code to listen to a presentation. I would prefer a more automatic method"*) in question set four (a=0.239), Item 24 *("The system is fun to use")* in question set five (a= 0.528), and Item 36 (*"I make a lot of response errors with the system (i.e., wrongly retained information)"*) in question set eight (a = 0.546).

Table 7.19 shows the questionnaire items with significant effects. A series of 2 x 2 ANOVAs, taking each questionnaire item as dependent variable and order of

presentation, type of ECA, and type of content as independent, showed significant effects of order of presentation on the following questionnaire items:

- Item 1 (*"The information task is too complex")* (F (1, 24) = 9.422; p < .01), and a significant interaction between the type of content and the type of ECA (see Figure 7.10) (F (1, 24) = 9.422; p < .01)
- Item 3 *("The process of extracting information from the system is difficult to learn")* (F (1, 24) = 8.075; p < .01), and a significant interaction between the type of ECA, and the type of content (see Figure 7.10) (F (1, 24) = 8.075 ; p < .01)

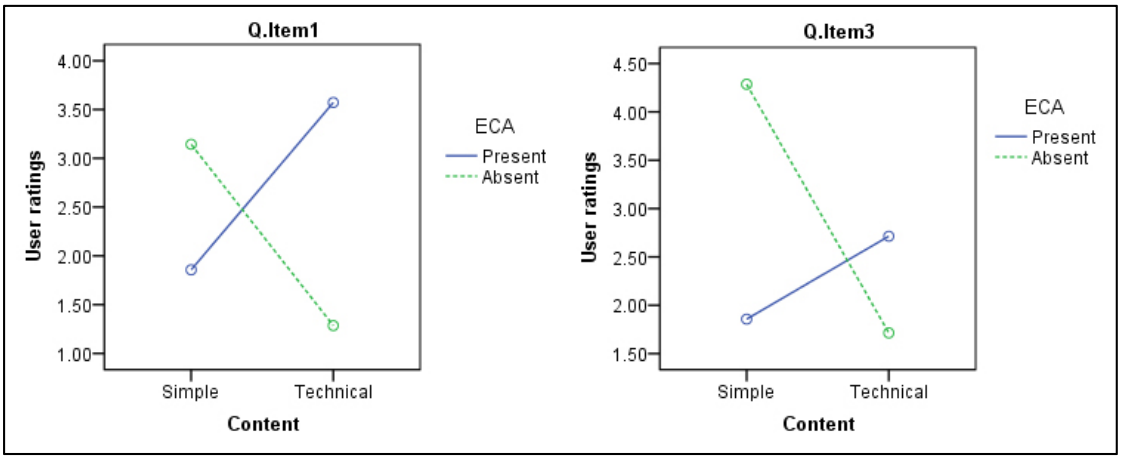


Figure 7.10: The interactions of ratings (Items 1, 3) for ECA and type of content

- Item 6 *("The completion of the information task requires too much self-organization")* (F (1, 24) = 5.481; p < .05), and a significant interaction between the type of content and the type of ECA (see Figure 7.11) (F (1, 24) = 5.481; p < .05) and finally,
- Item 26 (*"It's hard to learn any of the information presented by the system"*) (F (1, 24) = 7.032; p < .05) and a significant interaction between the type of content and the type of ECA (see Figure 7.11) (F (1, 24) = 7.032; p < .05)

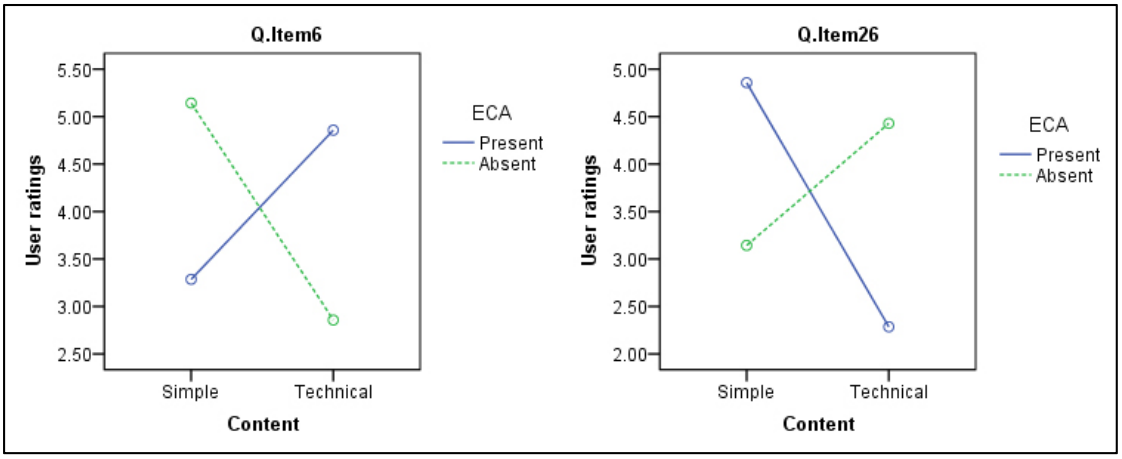


Figure 7.11: The interactions of ratings (Items 6, 26) for ECA and type of content

With regard to Item 1 (*"The information task is too complex"*), Table 7.19 shows that participants in the second order, thought that the information task was more complex (mean T/S = 3.3) than participants in the first order (mean S/T = 1.5). This is obviously due to the variation of the content across the two orders. Participants using the system with the ECA in the first order, experienced the simple content first (and rated the item lower) and then the technical content (and rated the item higher). Participants using the system without the ECA experienced the content vice versa and gave opposite ratings (i.e., lower for the first order, and higher for the second). The significant interaction between the type of ECA and the content for this item was further analysed using simple main effect analysis. It revealed that the variation of the content significantly influenced how the participants perceived the complexity of the information task with both the system with the ECA ($F(1, 24) = 4.342$; $p < .05$) and the system without the ECA ($F(1, 24) = 5.095$; $p < .05$). These two effects reveal an interesting pattern. Participants who used the system without the ECA thought that the technical content was less complicated (mean Technical = 1.2) than the simple content (mean Simple = 3.1). Then, participants who used the system with the ECA thought that the complexity of the task is lower with the simple content (mean Simple = 1.8) than with the technical content (mean Technical = 3.5).

		ECA			
Item	**Order of presentation**	**Present**	**Absent**	**AVG**	**Std. Deviation**
Item 1	S/T	1.8	1.2	1.5	1.0
	T/S	3.5	3.1	3.3	1.8
Item 3	S/T	1.8	1.7	1.7	1.3
	T/S	2.7	4.2	3.5	1.8
Item 6	S/T	3.2	2.8	3.0	2.3
	T/S	4.8	5.1	5.0	1.8
Item 26	S/T	4.8	4.4	4.6	2.0
	T/S	2.2	3.1	2.7	1.6

Table 7.19: Workload questionnaire items with significant order of presentation effects

The technical content may have seemed easier with the system without the ECA because of the text and voice used as output modalities. Then, because the presence of an ECA makes an interface more user friendly and the simple content was simple enough to understand with or without the ECA, it may have seemed to participants that the complexity of the task is lower with the system with the ECA than with the system without the ECA.

Then, although participants thought that the task is less complicated when experiencing the technical content with the system without the ECA, they had different views about its learnability. The descriptive statistics for Item 26 (*"It's hard to learn any of the information presented by the system"*) (see Table 7.19) show that participants in the second order thought that it is significantly easier to learn the information presented by the systems (mean T/S = 2.7) than the participants in the first order (mean S/T = 4.6). As in the previous questionnaire items, this effect is due to the variation of content across the two orders. A simple main effect analysis on the interaction between the order of presentation and the type of ECA, showed that the variation of the content significantly affected the participants using the system with the ECA ($F(1, 24) = 6.251$; $p < .05$) but not the participants using the system without the ECA. Participants thought that it is

significantly easier to learn the technical content (mean Technical = 2.2) with the system with the ECA than the simple content (mean Simple = 4.8). Therefore, I argue that the system with the ECA not only provides a more consistent method of presentation (see "Performance Measures"), but it also has the potential to enhance information retention of technical cultural content. As discussed later in the comments, participants felt that both contents were difficult to memorize, which explains the overall high ratings for this item with both systems. However, it seems that participants may have felt that the system with the ECA renders the interaction smoother thus making it easier for them to retain such information. This provides some evidence that supports my enhanced retention of complex information hypothesis (see *H8:* "*The enhanced retention of complex information"*). However as the objective measures failed to produce any significant results, it is hardly possible to draw any conclusions solely based on subjective evidence.

In relation to Item 6 *("The completion of the information task requires too much self-organization")*, Table 7.19 reveals that participants in the second order thought that the completion of the task requires more self-organization (mean T/S = 5.0), than the participants in the first order (mean S/T = 3.0). As in the previous questionnaire items, the significant interaction between the type of content and the type of ECA, shows that participants rated the ECAs across the two content conditions differently. However, the simple main effects analysis failed to reach conventional significance levels for either the system with the ECA or the system without the ECA. Therefore, this interaction can be best summarized as follows: Table 7.19 shows that participants thought they need more self-organization to complete the technical task with the system with the ECA (mean ECA-present = 4.8), than with the system without the ECA (mean ECA-absent = 2.8). They also thought, that they need less self-organization when they experienced the simple content with the system with the ECA (mean ECA-present = 3.2), than with the system without the ECA (mean ECA-absent = 5.1). A possible explanation is that the text used by the system without the ECA in the technical presentations was more natural for participants to read than watching an ECA on the screen giving information acting almost, but not perfectly, like an actual human being.

Curiously, the participants' retention performances do not follow the findings reported above. One would expect that since the system without the ECA renders the technical task less difficult and with less self-organization requirements, it would translate to enhanced retention performances with the system without the ECA. Then, if participants thought that it is easier to learn the technical information with the system with the ECA than with the system without the ECA their motivation should have resulted to enhanced retention performances. However, as it can be seen from Table 7.17, participants' retention performances when experiencing the technical content were similar with both systems (mean ECA-absent = 21.8 vs. mean ECA-present = 21.1) and improved when experiencing the simple content (mean ECA-present = 25.1 vs. mean ECA-absent = 36.8) with the system without the ECA. A possible explanation is that regardless of how the participants perceived the technical task with the system with the ECA, the modalities it used impacted their ability to memorise the content equally, as the modalities used by the system without the ECA.

Finally, for the significant effect of order on Item 3 *("The process of extracting information from the system is difficult to learn")* as Table 7.19 shows, participants in the second order, thought that the process of extracting information from the systems was more difficult to learn (mean T/S = 3.5), than the participants in the first order (mean S/T = 1.7). The significant interaction between type of content and type of ECA was further analysed using simple main effect analysis. It showed that the variation of content across the two orders significantly influenced how the participants perceived the difficulty of learning how to extract information from the system without the ECA ($F(1, 24) = 9.084$; $p < .05$) but not from the system with the ECA. Table 7.19, shows that participants rated the system with the ECA more consistently between the two content conditions (mean Simple = 1.8 vs. mean Technical = 2.7), than the system without the ECA (mean Technical = 1.7 vs. mean Simple = 4.2). Therefore, I can safely say that the variation of the content did not impact on how the participants perceived the learnability of the process when using the system with the ECA. Most likely, the modalities used by the system with the ECA made it easier for participants to learn how to extract information from the system. On the other hand, the variation of the content impacted how the participants perceived the process of extracting information with the system without

the ECA. The modalities used by the system without the ECA made it easier for participants to extract the information with the technical content, than the simple content.

Difficulty Ratings

A 2 x 2 ANOVA taking ratings as a dependent variable and the type of ECA and order of presentation as an independent, did not produce any significant effects for either the type of ECA or type of content. Participants on average rated the difficulty of the presentations higher with the system with the ECA, than with the system without the ECA (Table 7.20). Therefore, there is an indication that the participants perceived the way the content (simple or technical) was presented by the system without the ECA as less difficult, than when it was presented by the system with the ECA. However as discussed earlier (see the discussion for Item 3 of the Workload Questionnaire), while participants perceived the presentations with the system with the ECA as more difficult to understand, the modalities used resulted in more consistent retention performance as they are not affected by the variation of the content.

	ECA Present (n = 14)	**ECA Absent (n = 14)**
Mean	3.8	3.0
Std. Deviation	1.0	1.2

Table 7.20: Means for difficulty ratings

On the other hand, although participants perceived the presentations with the system without the ECA as less difficult, their retention performance was affected by the variation of the content. Participants performed better when they experienced the technical content with the system without the ECA than with the simple content.

Comments

Participants were asked to comment freely on each of the systems. The approach I used in the analysis of the collected interview data is discussed in Chapter 5 (see

§5.2.1.2). I present my findings below, grouped into relevant topics for simplicity. I discuss the corroborated patterns/comments first, and then the uncorroborated patterns/comments.

ECA Design:

1) Certain features of the avatar can be improved. These include: (Corroborated pattern by 2 out of 7 participants)

- Decrease the rate of the ECA's speech to make memorization easier
- Better body gestures
- More natural voice to avoid the speech discrepancies.

This corroborated pattern, suggests a number of improvements to the design of the ECA that, if implemented correctly, they could make memorization of the content easier. Participants did not have any comments about the photorealism of the ECA, which leads me to assume that it was acceptable, and focused only on the improvement of the ECA's behaviours and voice.

Multimodal Content Design:

1) The content is difficult to comprehend and memorize for most users. This could be because of the nature of the content, as it was suggested by a participant. A content of more historical value and without so many dates (as opposed to information about the construction of the churches), could be of more interest to the users. (Corroborated pattern by 5 out of 7 participants)

This pattern was corroborated by a number of participants. It shows that a different type of content (i.e., of a more historical value without so many dates) would be of more interest to the users. It also provides a possible explanation why participants scored overall low in the retention tests using both systems. A more personalised content to the preferences of participants may reveal stronger differences between the two systems (ECA-present and ECA-absent).

2) Shorter sentences (Uncorroborated pattern)

This pattern reflects the views of some participants, and although uncorroborated, it provides an interesting improvement that could be made to the content. Shorter sentences may enable easier memorization of the content.

Applications Design:

1) New features

- A zoom-in option to better see the artefacts for which the system is providing information (Corroborated pattern by 2 out of 7 participants)
- A pause button to pause the presentation on demand. (Corroborated comment by 1 out of 7 participants)

The above corroborated pattern and comment reveal features that should be added to the systems. Participants requested a pause button and a zoom option for the artefacts the system is providing information about. Although the ECA pointed to the artefacts, the resolution of the background images is low. The zoom feature will most likely make comprehension of the narrated content easier.

- A choice of repeating certain parts of the presentation. (Uncorroborated comment)

This comment though uncorroborated shows an interesting new feature that should be added to the systems. Participants requested a choice to repeat certain parts of the presentation. Although this would bias the retention performances (as some participants would use it and some others not), participants could be made aware of this feature and be allowed to use it only once.

2) Improvements in the existing design

- More multimedia elements (images and videos) to accompany the existing content. (Uncorroborated comment)
- A number of participants had problems photographing the QR-Code. (Corroborated comment by 1 out of 7 participants and also from the participants' observations)

The comments above require improvements in the existing design of the systems. In particular, participants requested more multimedia elements (images and if possible videos), and to improve the process of photographing the QR-Codes in the locations. According to the observations made by the experimenter, participants experienced two types of problems with the QR-Codes: a) some of the participants had problems photographing a QR-Code, as the lab was too bright. This can be solved by adding a higher resolution camera to the device. b) When the ECA was present, the experimenter noticed that the "click" sound of the system's camera did not work consistently. This was to be expected, as the UMPC device had to process the graphics of the avatar, in addition to the video stream needed to photograph the QR-Codes. The latest generation of UMPC devices offers significantly more processing and graphics power, than the device I used in my experiments and hence, I do not anticipate these problems to occur again.

Although the collected subjective and objective data would have been different in the real castle of Monemvasia, I believe that they provide a strong indication of how participants would perform and perceive the systems under real mobile conditions. The ECA would have still been perceived as a more consistent system in providing information of variable difficulty. While, the external environmental conditions (e.g., sunlight, noise) would have affected the ability of the ECA to effectively provide information, the latest generation of UMPC devices provides bright screens (visible even under direct sunlight conditions) and high-definition audio that would minimize these issues.

EXPERIMENT 3

Given that the routes chosen in the main study were all of average difficulty, there was a need for an additional experiment to examine the possible impact of the presence of a multimodal ECA on the user's ability to navigate routes of different complexity. In particular, I was interested in investigating the possible impact of the presence of an ECA on the participants' navigation performance in terms of time to complete a route and frequency of navigational errors and its effect on the subjective perception of the cognitive workload. My initial hypotheses were:

H9: An ECA enhances the participant's ability to navigate a complex route, but it does not have any effects on a simpler route.

H10: The perceived cognitive workload is lower in the system with the ECA than the system without one.

7.9 Overview

This experiment was performed with a methodology analogous to the previous experiments. It was designed to evaluate the impact of the presence of a multimodal ECA on the cognitive accessibility of a mobile tour guide system, providing navigation instructions in routes of varying difficulty. It manipulated the ECA's visual presence (present vs. absent), the difficulty of the route (simple vs. complex) and order of task (simple then complex route vs. vice versa). As in the previous experiments, it was conducted in the lab, but, instead of panoramic pictures, it used short high-resolution video clips (see Figure 7.12). The reason for this was the high cost of producing the panoramic scenes needed for the complex route. I found that video clips are cheaper to produce and easier to correct (e.g., when a clip does not show clearly the required landmarks).

Figure 7.12: One of the two interactive video applications

The video clips were assembled into two interactive applications representing in detail the two routes (simple and complex) participants had to follow in the real castle of Monemvasia. At each video-clip, the experimenter asked the participant what s/he would do at the particular point if s/he was in the real-castle of Monemvasia. To answer the question, the participant had to consider the provided visuals (i.e., landmarks as they appeared on the video clips, and the ECA gestures), and/or audio instructions delivered by the systems.

7.10 Design

In this section I report the design of experiment three. First, I give an overview of the participants and the software/hardware they used. Then, I report the task participants were asked to complete and the conditions under they completed it.

7.10.1 Participants

Two separate groups of nine participants (both males and females) from a variety of age groups took part in this experiment (see Table 7.21 and D.3.1 in Appendix D for the full participant details). Because of the simplicity of the system, I asked only one user to review the system and instruments of research. A series of minor updates were made based on his feedback. Again, in order to avoid over-familiarity with the selected routes, no participants were either a local resident or had visited the area before. Participants had a variety of academic and mobile-computer backgrounds and were native English speakers.

Order of task	Participants	Age (Mean)	Std. Deviation	Gender (M/F)
Simple vs. Complex	9 Participants	21.5	1.6	7/2
Complex vs. Simple	9 Participants	22.8	3.4	8/1

Table 7.21: Table of participants in experiment three

7.10.2 Software and Equipment

A stripped-down version of the main experimental systems was used in this study. In particular, the systems were identical to the main ones, but capable of only providing navigation instructions (based on photographs of landmarks) (see Figure 7.13). The navigation instructions of the same type (simple or complex) were of the same length and complexity. All the presentations about locations of the castle were removed from both systems.

Figure 7.13: The system with the ECA (left side) and the system without the ECA (right side)

7.10.3 Task

Participants were told that the purpose of this experiment was to investigate different systems for path finding with respect to their impact on their ability to find their way in the castle. They were asked to navigate two different routes (simple route then complex or vice versa) visiting a number of locations in turn, using the system A (i.e., the visual agent) on one occasion, and the system B (i.e., the non-visual agent) on the other. The chosen routes included 13-17 waypoints (represented by short-video clips) and took 15 – 25 minutes to complete. In addition, participants were told that an experimenter would be present in the lab to monitor their navigation behaviour, measure the number of times they got lost and provide help if necessary. After each route, participants were asked to indicate on a 5-point scale which of the two systems they found most useful and to give reasons for their preference. As the question was asked verbally, I kept the scale to 5-point to reduce the load to the participants' working memory. The five point scale is simpler and

better understood than higher scales, and it offers a good range of options to capture the usefulness of the systems. Finally, they were asked to fill-in a seven-point Likert (1932) format questionnaire measuring the cognitive workload invested in navigating the routes with each of the systems.

7.10.4 Conditions

Only two within-subjects variables were manipulated: type of agent (ECA-present vs. ECA-absent) and route complexity (simple vs. complex). The variables order of task (simple then complex route vs. vice versa) was manipulated between-subjects (see Table 7.22). Participants were randomly assigned to the four experimental conditions: 1) Present ECA with the simple route vs. absent ECA with the complex route or 2) Present ECA with the complex route vs. absent ECA with the simple route.

Participants (n = 18)	**ECA Present**	**ECA Absent**
1 – 9 Participants	**Simple Route** Time/Errors/Preference Rating/Workload Questionnaire	**Complex Route** Time/Errors/Preference Rating/Workload Questionnaire
10 – 18 Participants	**Complex Route** Time/Errors/Preference Rating/Workload Questionnaire	**Simple Route** Time/Errors/Preference Rating/Workload Questionnaire

Table 7.22: Experimental design of experiment three

7.11 Measures and Methods

Both objective and subjective measures were used in this experiment. The three objective measures were:

- Time taken to navigate each route
- Number of times the participants got lost on each of the routes, and the
- Preference ratings.

"Lost", was defined as a wrong answer to the question about what the participant would do at the particular video clip. A wrong answer was assumed to divert the participant from the planned route, which required the experimenter to intervene in order to get him/her back onto the route. The primary subjective variable was the participants' responses to the individual items in the workload questionnaire. These data were collected through direct observation, device logs and a questionnaire. In particular, the experimenter made notes on each participant's navigation behaviour and measured the number of times they got lost. The time taken to navigate the routes was measured through each system's internal chronometer functionality. The questionnaire was divided into two parts. The first part, asked participants to indicate whether (yes/no) they saw specific objects on each route (e.g., a path). This was measured as an indication of the effectiveness of each system in directing the participant's attention to landmarks in his/her physical environment. The second part, used a seven-point Likert (1932) scale (1=strongly disagree, 7=strongly agree) format and measured the perceived cognitive workload as an indication of how the participants felt about using the systems to navigate the specified routes. This part of the questionnaire was based on the Simplex Two theory (discussed in §3.2.2 of Chapter 3) and included items such as complexity of navigation, user frustration, etc.

7.12 Results and Discussion

The following section discusses the results of the third experiment. It begins with the quantitative measures, i.e., "the time take to complete a tour", and "the frequency of getting lost". It then continues with a discussion of the results of the subjective assessment. There are results from the "Object Recognition" and the "Workload" questionnaires, the usefulness ratings and the comments participants had after the completion of the task.

Performance Measures

A 2 x 2 ANOVA taking type of ECA, and order of task as independent variables and time (in seconds) (see Table 7.23) as a dependent variable, showed no significant effects for type of ECA or order of task.

	ECA Present (n = 18)	ECA Absent (n = 18)
Mean	1155.8	1019.4
Std. Deviation	315.5	218.0

Table 7.23: Mean times to complete a tour

There was a significant interaction between the type of ECA and order of task ($F(1, 32) = 11.940$; $p < .01$). This interaction effect was analysed using simple main effects analysis. The variation of order of task significantly influenced the time performance of participants using the system without the ECA ($F(1, 32) = 13.213$; $p < .01$), but not the system with the ECA. The descriptive statistics (see Table 7.24) suggest that the participants spent significantly more time with the system without the ECA in the first order (mean time S/C = 1217.5 sec), than in the second order (mean time C/S = 821.3 sec). However, there was no such a difference in the participants using the system with the ECA.

ECA	S/C (n=9)	Std. Deviation	C/S (n=9)	Std. Deviation
Present	1087.6	423.2	1224.1	148.3
Absent	1217.5	96.8	821.3	57.6

Table 7.24: Time performance as a function of ECA and order of task

Participants using the system with the ECA, spent slightly more time on the complex route (mean Complex = 1224.1 sec), than in the simple route (mean Simple = 1087.6 sec). One possible explanation is that participants focused more their attention on the additional modalities (other than speech) provided by the two systems (i.e., text or gestures), to disambiguate the navigation instructions. In the system without the ECA, participants commented (see comments), that they had to be fast in reading the text especially when the instruction was too long. If they would fail to read the text they would become unsure what step to take.

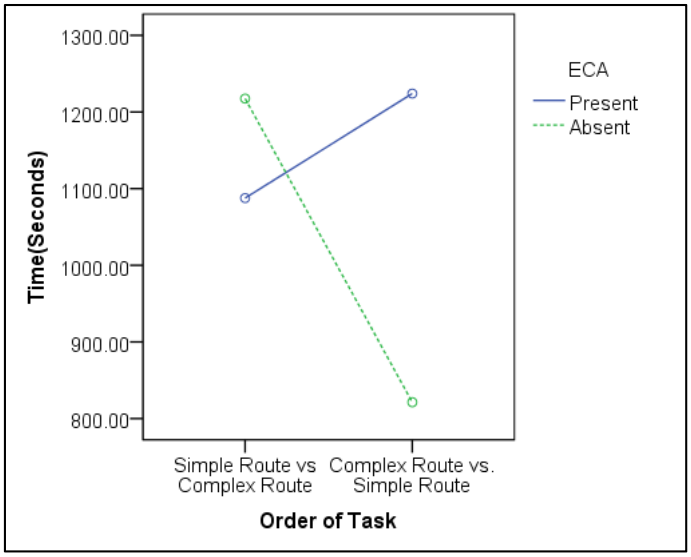


Figure 7.14: The interaction of time for ECA and order of task

The gestures used in the system with the ECA, did not have such a negative impact on the participants. As they commented, gestures were accurate enough to help them in better understanding the speech and hence to take easier decisions on where to go.

With respect to the number of times participants got lost, a 2 x 2 ANOVA taking order of task and type of ECA as independent variables and frequency of getting lost as a dependent showed no significant effects. No interactions were found either. However participants got lost less often with the system with the ECA than with the system without the ECA. In particular, they got lost on average 2.5 times when using the visual agent system compared with 3 times when using the non-visual system (see Table 7.25).

	ECA Present (n = 18)	**ECA Absent (n = 18)**
Mean	2.5	3
Std. Deviation	1.58	1.64

Table 7.25: Summary of means of getting lost from D.3.3

A possible explanation is similar to the one given for time performances. If participants found the visual agent's gestures more helpful in disambiguating the provided navigation instructions than the text used by the system without the ECA,

then it is to be expected that they got lost fewer times with the visual agent, than with the non-visual agent. The performance data invalidated my original hypothesis (see H9). The presence of the ECA does not enhance the participant's navigational ability with a complex route. In fact, it has no measurable impact on either the complex or the simple route. On the other hand, the variation of the route affected the participants using the system without the ECA. Their performance was better with the simple route than with the complex route. With regards to the frequency of getting lost, the data show that the ECA can enhance navigation performance, but the difference between the two means is not significant. Hence, it is safe to say that an ECA does not enhance navigation performance, but it provides a more consistent presentation method for routes of varying difficulty, than a system without such an artefact on the interface.

Subjective Assessment

Object Recognition Questionnaire

The following table (see Table 7.26 and Table D.3.5 in Appendix D for more details), shows the total number of objects participants recognised, and the objects that missed across the two orders with each of the systems. A chi-squared test returned no significant results within or between any of the conditions. Therefore, it can be concluded that the participants recognized effectively the majority of the objects the systems provided information about.

Order of task	ECA Present (n = 18)	ECA Absent (n = 18)
Simple/Complex	49/5	48/6
Complex/Simple	43/11	44/10
Total (Y/N)	92/16	92/16

Table 7.26: The physical object (Y/N) recognition results

Workload Questionnaire

As with the questionnaires in the previous experiments, the workload questionnaire for this experiment consists of nine questions-sets (see Table D.3.6 in

Appendix D for more details), each asking questions reflecting the requirements of the Simplex II model. The first set is consisted of six items, and it was designed to evaluate a number of aspects of the navigation task i.e., complexity, learnability, consistency and self-organization aspects. The second set examined how the participants perceived the output modalities of the prototypes (visual, auditory and textual) in terms of sensory, satisfaction and understanding, and it includes four items (e.g., understanding of the navigation instructions, visibility of the device screen etc.). The third set evaluated how the participants felt about the feedback given by the prototypes, in terms of timing, relevance and memory requirements (e.g., timing of the system output, support to locate landmarks, etc.). This set includes six items. The fourth set of questions assessed the demands the prototypes placed in the participant's working memory, and it includes four items (e.g., remind participants when they get off the route, the utility of the manual method to get navigation instructions, etc.). The fifth set evaluated the emotional impact of the prototypes (e.g., fun, annoying, etc.) and it includes four items. The sixth question set evaluated the impact of the prototypes on the participant's long-term memory in terms of the task learnability and relation to their existing knowledge. This set includes four items (e.g., ease of learning of a route, the relation of a route to a participant's personal interests, etc.). The seventh set consists of four items, and it was designed to assess how effective participants could access the underlying structure of each route (e.g., simplicity of a route, how the structure of a route is presented, etc.). The eighth set evaluated how participants felt about the rationality of their responses, and the degree of support they received during their responses. This set includes four items (e.g., allowances for response errors, number of wrong navigation decisions etc.). The ninth and final set of items evaluated how participants felt about their output responses and the support they received in order to respond appropriately. It consists of four items (e.g., ease of finding landmarks, support provided to learn the navigation instructions, etc.)

As can be seen in Table 7.27, most alphas range from acceptable to good, with a few exceptions. A close examination of the "Item-Total" statistics reveals the problematic questionnaire items of each group. In particular: Item 18 ("*I find it difficult to remember that I have to tap a button to get the next instruction. I would prefer a more automatic method"*) in group 4 (a = 0.729), Item 24 ("*The system is*

fun to use") in group 6 (a = 0.888), and Item 32 *("A simpler route would have been more enjoyable and easier to actually learn"*) in group 7 (a = 0.763). No other problematic items were found. These items will be revisited or removed from future versions of the questionnaire.

Q.Set	Alpha
1	0.840
2	0.732
3	0.817
4	0.685
5	0.081
6	0.577
7	0.653
8	0.534
9	0.651

Table 7.27: Cronbach alphas of the workload questionnaire

I conducted a series of 2 x 2 ANOVAs, taking each questionnaire item as a dependent variable, the type of ECA, type of route, and order of task as independent variables. I found an effect for ECA on Item 10 (*"It is difficult to make sense of the instructions used by the system")* ($F(1, 32) = 5.434$; $p < .05$) and an effect of order of task on the following items:

- Item 11 (*"The navigation instructions provided by the system is poorly presented (too brief or too long)")* $F(1, 32) = 6.516$; $p < .05$) and
- Item 27 (*"The routes provided by the system should better relate to my personal interests"*) $F(1, 32) = 6.527$; $p < .05$.

There was also a significant interaction between the type of ECA and the type of route for the following items (see Figure 7.15):

- Item 11 (*"The navigation instructions provided by the system is poorly presented (too brief or too long)")* ($F(1, 32) = 6.527$; $p < .05$) and

- Item 27(*"The routes provided by the system should better relate to my personal interests"*) (F (1, 32) = 6.527; p < .05)

No other significant effects or interactions for any other of the questionnaire items were found.

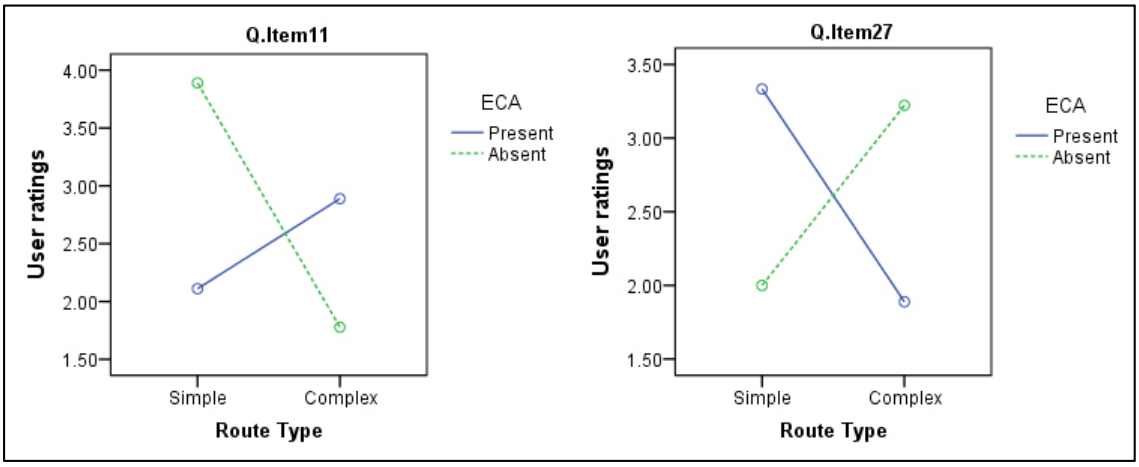


Figure 7.15: The interactions of ratings (Items 11, 27) for ECA and type of route

The results for Item 10 (*"It is difficult to make sense of the instructions used by the system")*, suggest that participants experienced the difficulty of the navigation instructions differently with ECAs of the same type. Participants using the system without the ECA gave higher difficulty ratings for the navigation instructions (mean ECA-absent = 3.4) than the participants using the system with the ECA (mean ECA-present = 2.1). This finding is in line with my hypothesis (see H10) and the comments made by the participants (see comments below). The ECA augmented the acoustical information with relevant gestures to show the participants where to go. The combination of relevant body gestures with speech made it easier for the participants to make sense of the instructions given by the ECA (and to take better navigation decisions) than the system without the ECA that used only voice and text.

With regard to Item 11 (*"The navigation instructions provided by the system is poorly presented (too brief or too long)"*), as can be derived from the descriptive statistics in Table 7.28, the participants in the S/C condition rated the navigation

instructions of both systems better (mean S/C = 1.9), than the participants in the C/S condition (mean C/S = 3.3). A closer inspection of the descriptive data reveals, that participants rated the system with the ECA similarly across the two orders (mean S/C = 2.1 vs. mean C/S = 2.8), but not the system without the ECA (mean C/S = 1.7 vs. mean S/C = 3.8). This shows that the ECA presented navigation instructions of variable difficulty in a more consistent manner than the system without the ECA. This consistency can be attributed to the gestures used by the ECA in both types of routes. The gestures were well synchronised with the speech and designed in a way to augment the instructions provided. On the other hand, the text used by the system without the ECA though synchronised with the speech, made it difficult for participants to read the instructions (especially when an instruction was too long).

		ECA			
Item	**Order of task**	**Present**	**Absent**	**AVG**	**Std. Deviation**
Item 11	S/C	2.1	1.7	1.9	1.2
	C/S	2.8	3.8	3.3	2.0
Item 27	S/C	3.3	3.2	3.2	1.6
	C/S	1.8	2	1.9	1.3

Table 7.28: Workload questionnaire items with significant order of task effects

Then, I analysed the ECA by type of route interaction using simple main effect analysis. I found that the variation of route significantly influenced how the participants perceived the instructions when presented by the system without the ECA ($F(1, 32) = 6.959$; $p < .05$), but not by the system with the ECA. In particular, the participants perceived the instructions provided by the system without the ECA as more poorly presented in the simple route (mean Simple = 3.8), than in the complex route (mean Complex = 1.7). The most likely explanation for this interaction is that the navigation instructions provided by the system without the ECA in the simple route were very brief and did not provide enough information to the participants to aid their navigation decisions. Finally, the difference between the system with the ECA, across the two types of routes was minimal (mean Simple = 2.1 vs. mean Complex = 2.8), which validates the discussion above about the

system with the ECA presenting navigation instructions in a more consistent manner than the system without the ECA.

With regard to item 27 (*"The routes provided by the system should better relate to my personal interests"*), Table 7.28 shows that participants in the first order giving higher ratings (mean S/C = 3.2) for both systems than the participants in the second order (mean C/S = 1.9). The most likely cause factor of this effect is the variation of the route complexity. The significant interaction between the type of ECA and the type of route suggests that participants rated the degree of interest of each route differently between the two order conditions and for the same type of ECAs. In particular, the simple route (see Table 7.28) was perceived as less interesting than the complex route (mean Simple = 3.3 vs. mean Complex = 1.8) when each route was presented by the system with the ECA. Then, the complex route was perceived as less interesting than the simple route (mean Complex = 3.2 vs. mean Simple = 2) when each route was presented by the system without the ECA. However, a simple main effect analysis showed that none of these differences reached statistical significance levels. This finding is in line with the significant effect of the type of ECA on Item 10 *("It is difficult to make sense of the instructions used by the system")* discussed above. Participants found the complex route when they experienced it with the system with the ECA, to be not only less difficult than with the system without the ECA, but also more related to their personal interests. In contrast, participants perceived the simple route when presented by the system without the ECA as more interesting and less difficult than the complex route.

Usefulness Ratings

Table 7.29 shows a summary of the ratings (in the 5-point scale) that participants gave for each of the systems. A 2 x 2 ANOVA taking the type of ECA and order of task as independent variables and the mean ratings as dependent variable produced a significant effect of order of task ($F(1, 32) = 7.040$; $p < .05$). This finding follows the general effects found on the questionnaire items discussed above.

Order of task	ECA (n=18)	Mean	Std. Deviation
Simple vs. Complex	Present	2.4	1.8
	Absent	2	1.6
Complex vs. Simple	Present	3.8	0.92
	Absent	3	0.70

Table 7.29: Summary of usefulness ratings from D.3.4

The usefulness of the ECA was rated higher in the complex route (mean Complex = 3.8) than in the simple route (mean Simple = 2.4). On the contrary, the usefulness of the system without the ECA was rated lower (mean Complex = 2) in the complex route than in the simple route (mean Simple = 3). One explanation is that because the simple route was simple enough to navigate without any help, participants considered the presence of the ECA unnecessary.

It is most likely that the pictures of the landmarks used as a background by the systems, were enough for effective navigation of the simple route. The system without the ECA because the text instructions that accompanied the voice were difficult for participants to read (see comments below), participants rated the usefulness of the system lower in the complex route, than in the simple route.

Comments

After rating the usefulness of each system for providing navigation instructions, participants had to justify their ratings for each of the systems. I followed the same approach (see §5.2.1.2 of Chapter 5) as in the previous two experiments to analyse the gathered data. Four participants were excluded from the analysis as they had no comments. Below, I present and discuss my findings in more detail.

ECA Design:

1) It is better to have the ECA, because it shows you where to go with relevant body movements (Corroborated pattern by 6 out of 7 participants)

This pattern was corroborated by six out of seven participants, and explains why participants thought the utility of the ECA is better in providing navigation instructions than the system without the ECA. In particular, the system with the ECA was seen as more useful, as it uses relevant body gestures to show the participants where to go.

2) The system with the ECA was more user-friendly, but had no effect on the instructions provided. (Uncorroborated comment)
3) The ECA should include text along with voice and gestures to give directions. (Uncorroborated comment)

The first uncorroborated comment, though contradictory to what most participants believe shows that the ECA could render the interface more user friendly, but it has no impact on the comprehension of the navigation instructions provided by the systems. Although the comment reflects the views of some participants, it was not corroborated and, hence, it remains speculative. The second uncorroborated comment reveals the possibility that the effectiveness of the ECA could be improved with the addition of text, along with voice and gestures. Although it is a possibility, it is most likely that the addition of so many modalities on an interface will overload participants and decrease the effectiveness of the ECA instead of improving it.

Subtitle/text Design:

1) The subtitles distracted as it was difficult to read, listen to the instructions and decide where to go at the same time. (Corroborated pattern by 3 out of 7 participants)
2) The voice that reads the subtitles was too fast (faster than the system with the ECA) (Corroborated pattern by 2 out of 7 participants)

Both of the corroborated patterns above explain why the participants thought the system without the ECA was less effective in providing navigation instructions than the system with the ECA. First, the voice used along with the text was too fast. Secondly, as there was no scrolling option in the window that displayed the text,

participants had to be quick readers. As they also had to listen at the same time and the text hampered their ability to take the correct navigation decisions.

3) The subtitles helped, as the user could look back to the text in case s/he would forget or could not understand an instruction. (Uncorroborated comment)

This comment, though uncorroborated reveals that for the participants who were fast readers, the text used by the system without the ECA was actually helpful. They could use to the text as reference in case they would forget or failed to understand a navigation instruction.

Videos:

1) It is hard to relate the systems with the videos. The videos should be clearer (Corroborated pattern by 1 out of 7 participants)

As the videos were recorded months after the pictures of the landmarks displayed by the systems, it was difficult for the participants to relate the content of some navigation instructions with the videos in order to decide where to go. In addition, the quality of the videos was not very high, which made it even more difficult to relate the system content with locations displayed in the videos.

Multimodal Content Design:

1) The complex route provided instructions that the systems (ECA or subtitles) spoke very fast. That makes it difficult to remember what has actually been said. (Corroborated pattern by 2 out of 7 participants)

This corroborated pattern shows that the participant thought, the instructions provided by both systems in the complex route, were too difficult to remember. However, as discussed in the analysis of the Workload questionnaire, participants thought it was less difficult to remember the complex instructions with the system with the ECA, than with the system without the ECA.

2) There is no difference between the system with the ECA and the system without the ECA. However the system that provides the simple instructions (regardless of the presentation method) is easier to understand. (Uncorroborated pattern)

This uncorroborated pattern reveals an ECA-zero effect. The manipulation of the type of ECA (present vs. absent) did not have any effect on the participants' understanding of the navigation instructions in both types of routes. However, the complexity of the route (irrespective of the type of ECA) impacts the participants' understanding of the navigation instructions. The instructions provided in the simple route, were easier to understand than in the complex route.

Although the collected subjective and objective data do not accurately represent a real outdoor environment, they provide a strong indication of what is most likely to happen when the system would be tested in the real castle of Monemvasia. I believe that although the visibility and acoustics of the content would have been affected by external environmental conditions (e.g. sun, noise etc.), the ECA would have still been more effective in enhancing the participants' ability to disambiguate the acoustic/visual instructions than the system without the ECA. Obviously the external environmental conditions, would have affected the ECA's ability to effectively give directions, but the latest generation of the UMPC devices offer very bright screens (visible even under direct sunlight conditions), and high definition audio that limits the impact of the environmental conditions.

7.13 False Positive Questionnaire Results

The three experiments discussed above used a cognitive accessibility questionnaire that was based on the Simplex II model (see §3.2.2 of Chapter 3) of human cognition, in order to capture the views of the participants. The Simplex II model postulates nine components of human cognition. Each questionnaire contains forty questions. Further details of the questionnaires are provided in appendix D.1.7, D.2.4 and D.3.6.

The questionnaires were completed for each experimental condition of each of the three experiments. Of course, there is a risk of false positives when analysing multiple conditions, so first, the number of significant results is compared against the estimated false positive result. For example, when testing forty (40) questionnaire items at the 5% level of significance, then two apparently significant results (40 x 0.05 = 2) would be expected by chance. However, significant results occurred much more frequently and at higher significance levels, as shown below.

In experiment one, I found five significant results for the variable order, with an average significance level of p = 0.0262. At this level of p, then only 1.048 significant results would be expected by chance i.e. approximately one. So, clearly, these five significant results cannot be dismissed as false positives. A simple perspective of the Simplex model might expect any significant results to cluster around the same putative modules, of which there are nine. This is clearly not the case, as Table 7.30 shows. The modules involved were: three, five, seven and nine; with only module three showing more than one significant item (Items 12 and 13). Then, I found nine significant results for the variable scenario, with an average significance level of p = 0.015222. At this level of p, then only 0.000761 (less than one) significant results would be expected by chance. So, clearly, these nine significant results too cannot be dismissed as false positives. In experiment two, I found four significant results for the variables order and interactions, with an average significance level of p = 0.014. At this level of p, then only 0.56 (less than one) significant results would be expected by chance. So, clearly, these four significant results too cannot be dismissed as false positives. Three of the four significant results were related to the same Simplex module (i.e., module one; Executive Functions). In experiment three, I found only one significant result for the variable ECA. This was significant at the 0.26 level. Unfortunately, one is the number of false positives to be expected at this level of significance, so this result should be ignored. Also, in this experiment, I found two significant results for the order and interaction variables, with an average significant level of p = 0.016, so this is still more than the expected false positive rate would be predicted to be found at p = 0.016 (< 1, 0.64).

Overall, the Simplex modules four (Working Memory), and eight (Output Responses) showed no significant results and two modules showed only one significant result each: module five (Emotions and Drives) and module seven (Mental Models). Module six (Long Term Memory) produced three significant results. The following modules each generated four significant results: Module one (Executive Functions), Module two (Perception/Input) and Module three (Feedback).

Exp.	Factor	Q. Item	P value	Simplex Modules	Notes
1	Order of systems	12	.000	Module three (Feedback Management)	Five significant results Mean p = 0.0262 40 x 0.0262 = 1.048
1	Order of systems	13	.038	Module three (Feedback Management)	
1	Order of systems	23	.042	Module five (Emotions and Drives)	
1	Order of systems	30	.033	Module seven (Long Term Memory)	
1	Order of systems	39	.018	Module nine (Output Sequence)	
1	Scenario	2	.039	Module one (Executive Functions)	Nine significant results Mean p = 0.015222 40 x 0.015222 = 0.60888
1	Scenario	7	.035	Module two (Perception/Input)	
1	Scenario	8	.000	Module two (Perception/Input)	
1	Scenario	10	.000	Module two (Perception/Input)	
1	Scenario	12	.004	Module three (Feedback Management)	
1	Scenario	25	.018	Module six (Long Term Memory)	
1	Scenario	28	.016	Module six (Long Term Memory)	
1	Scenario	38	.025	Module nine (Output Sequences)	
1	Scenario	40	.000	Module nine (Output Sequences)	

2	Order of presentation & interactions	1	.005	Module one (Executive Functions)	Four significant results Mean p = 0.014 40 x 0.014 = 0.56
2	Order of presentation & interactions	3	.009	Module one (Executive Functions)	
2	Order of presentation & interactions	6	.028	Module one (Executive Functions)	
2	Order of presentation & interactions	26	.014	Zone nine (Output Sequences)	
3	ECA	10	.026	Module two (Perception/Input)	One significant result Mean p = 0.026 40 x 0.026 = 1.04
3	Order of task	11	.016	Module three (Feedback Management)	Two significant results Mean p = .016 40 x .016 = 0.64
3	Order of task	27	.016	Module six (Long Term Memory)	

Table 7.30: Summary of significant results in the three experiments

7.13.1 Combining Estimates of Statistical Significance

In the first experiment, I found five significant results for the variable order with an average significance level of p = 0.0262. Using the natural logarithm equation, $X^2 = 2\ \Sigma\text{-}\ln P_i$, I calculated the overall estimates of statistical significance for the impact of each factor on the questionnaire responses of the participants in the three experiments[40]. The impact of this factor on the questionnaire responses was very highly significant X^2 (10) = 46.16, p < .001 (see Table 7.31). Then, I found nine significant results for the variable scenario, with an average significance level of p = 0.60888. The impact of this factor on the questionnaire responses was very highly significant X^2 (18) = 117.18, p < .001 (see Table 7.32).

[40] To calculate the natural logarithmic of each factor I used the Natural Logarithmic calculator at http://www.rapidtables.com/calc/math/Ln_Calc.htm

	P value	**-lnP_i**
1	.000	9.21
2	.038	3.27
3	.042	3.17
4	.033	3.41
5	.018	4.02
Total		23.08
Df = 2k (number of significant results) = 2x5 = 10		
X^2= 2 (23.08) = 46.16, p < .001[41]		

Table 7.31: Impact of order of systems in experiment one

	P value	**-lnP_i**
1	.003	5.81
2	.018	4.02
3	.000	9.21
4	.000	9.21
5	.000	9.21
6	.012	4.42
7	.022	3.82
8	.025	3.68
9	.000	9.21
Total		58.59
Df = 2k (number of significant results) = 2x9 = 18		
X^2= 2 (58.59) = 117.18, p < .001		

Table 7.32: Impact of scenario in experiment one

In experiment two, I found four significant results for the order and interaction variables, with an average significance level of p = 0.014. The impact of this factor on the questionnaire responses was very highly significant X^2 (8) = 35.72, p < .001 (see Table 7.33).

[41] To calculate the value of p I used the chi-square table at http://www.medcalc.org/manual/chi-square-table.php

	P value	**-lnP_i**
1	.005	5.30
2	.009	4.71
3	.028	3.58
4	.014	4.27
Total		17.86
Df = 2k (number of significant results) = 2x4 = 8		
X^2= 2 (17.86) = 35.72, p < .001		

Table 7.33: Impact of order of presentation and interaction in experiment two

In experiment three, I found three significant results, with an average significant level of p = 0.0193. The impact of this factor on the questionnaire responses was highly significant (see Table 7.34) X^2 (6) = 23.86, p < .01.

	P value	**-lnP_i**
1	0.026	3.65
2	0.016	4.14
3	0.016	4.14
Total		11.93
Df = 2k (number of significant results) = 2x3 = 6		
X^2= 2 (11.93) = 23.86, p < .01		

Table 7.34: Impact of order of task and ECA in experiment three

7.13.2 Discussion of Questionnaire Results

I evaluated the questionnaire data from experiments one, two and three from two different perspectives, but both essentially agree that the present results are of substantial statistical significance. First, I found that the numbers of significant results exceeded the estimated numbers of false positive results that would be expected by chance. Second, using the equation X^2= 2 Σ-lnP_i, I calculated overall estimates of statistical significance for the impact of each factor on the questionnaire responses of the participants in the three experiments. I found that for

the four factors, three were very highly significant ($p < .001$) and one factor was highly significant ($p < .01$).

Two meaningful themes emerge from the present analyses. First, the different factors appear to influence different items of the questionnaire. That is, probably, to be expected. Second, perhaps less expected, whilst a simple perspective of the Simplex theory might expect any significant results to cluster around the same putative Simplex modules, of which there are nine. This is clearly not the case, as Table 7.30 shows: module two (Perception/Input) had three significant items (Items 7, 8 and 10); module six (Long Term Memory) had two significant items (Items 25 and 28); module nine (Output Sequences) had two significant items (Items 38 and 40); modules one and three had one significant item each. Thus, it seems that either the Simplex modules are not validated or that the specific factors impacted specific subsets of each factor and not the overall factors themselves. Further work is needed to explore these findings further.

7.14 Conclusions

The first experiment examined the effect of an ECA compared with a text/voice system in helping users navigating routes in a simulated outdoor environment and extract personalised information from various attractions. In this study, participants using an ECA capable of multimodal input and output did not perform significantly differently in terms of enhanced navigation and information retention than the participants using a system with text and voice output. However, I found evidence that an ECA can be more effective in enhancing the participants' ability to navigate routes that required more complex navigation decisions than the system without the ECA. In terms of information retention, I found evidence that an ECA is a more consistent presentation method than a system without it. Given this consistency, if the modalities used by the ECA to present information are improved (e.g., by synchronizing the ECA's body language well with the content) then an ECA could potentially benefit retention performance instead of degrading it. One last important finding is that the objective performance is affected by the user's personal constraints. I observed that the participant's time constraints impacted the way they interacted with the systems and the amount of information they received. Although

in the simulated mobile environment, this factor did not impact significantly retention performance, in the real castle of Monemvasia it would most likely significantly affect what users retain from each presentation. Therefore, it is important for any guide system to adapt the quality and quantity of the presentations to accommodate both short-term (e.g., someone who is a passing-by visitor to the castle of Monemvasia) and long-term visitors (e.g., someone who has holidays for a few days in the castle).

With respect to subjective assessment, I found strong evidence that the degree of content personalization strongly affects how participants perceive the systems, tasks and the ECA. If the cultural content does not match the user's interests and background knowledge, their overall experience with the systems is highly dampened. Therefore, the primary focus of a system's designer should be to improve the accessibility of the content, with the improvement of the performance of the systems and/or the ECA to be a secondary design goal. Ideally, the content should be universally accessible to all users, regardless of their background knowledge and experience. This study, generated some ideas of how to achieve this goal (e.g., author content with limited use of historical dates), but clearly this issue needs further investigation.

Then, this experiment provides encouraging news on the use of panoramas as a tool for simulating an outdoor mobile environment in the lab. Although, I observed that some participants had problems synchronising their movements between the panoramas and the mobile device, they were all able to complete the assigned tasks mostly without prior training. If the interaction problems are solved and a treadmill is added to simulate walking from location to location, then the overall user's experience would come closer to that of visiting the real castle of Monemvasia.

Finally, I observed that although the participant's views on whether or not an improved ECA (in the calibre of the "Avatar" movie) would be more effective in providing information (cultural content or navigation instructions) were divided, there was not even a single participant that did not remember the story of the movie and was not impressed by the cutting-edge computer animation technology, used to create the characters. Therefore, I argue that as computer character animation

becomes a dominant form of movie-making, it will eventually lead to a greater user acceptance of ECA's on computer interfaces.

The finding of experiment one regarding the ECA presentation consistency is repeated in experiment two. In this study, I varied the content between the two ECA conditions more aggressively by exposing participants to simple and technical content. As in experiment one, I found that although there were no statistically significant differences between the type of ECA and type of content, participants who used the system with the ECA performed more consistently with content of varying difficulty, than those who used the system without the ECA. However, as discussed below an ECA (with text as an additional output modality) can positively impact the perception of technical content. This can potentially lead to enhanced retention of technical information.

The subjective assessment showed that the cognitive load of participants, with regards to the organization and implementation requirements of the information task (see Simplex II model in §3.2.2 of Chapter 3), was affected by the complexity of the content, and the type of system that presented it. The system without the ECA lowered the difficulty and self-organizational requirements of the technical task, but placed more demands to the participants in terms of its learnability. The system with the ECA, on the other hand, lowered the learnability of the technical task, but had a detrimental effect on the perception of its difficulty and self-organization. Although this finding can be generalised only with caution, the careful use of text as an additional output modality when presenting technical content in an ECA-based mobile guide system can be beneficial for users rather than degrading.

Then, the use of QR-Codes as a technique of content-tagging random locations in an outdoor attraction was received positively by the participants even under simulated conditions. Although some participants experienced issues with photographing accurately a QR-Code the first time, they all managed to complete a full tour with both systems with no prior training.

In the third experiment, the finding of consistency is again repeated. The ECA provided a more consistent presentation method for routes of varying difficulty than

the system without the ECA. Although, I found no statistically significant differences (possibly due to the simulated conditions), participants using the system with the ECA were lost overall fewer times than those who used the system without the ECA. In fact, based on the gathered qualitative data and direct observations of the participants, I argue that the usefulness of an ECA in a mobile guide system is stronger in the navigation task than the information task. Using relevant body language, the ECA can illustrate navigation instructions (e.g., when you come in front of the fork-path, turn left) that are impossible to illustrate using other methods (e.g., an animated arrow).

Qualitative results revealed what I had already observed in the lab. Participants perceived the navigation instructions as they were presented by the ECA as more clear and better presented than the system without the ECA. Furthermore, I found that the ECA affected how the participants perceived the complex route in relation to their personal interests, but not the simple route. Therefore, it is safe to say that although an ECA does not enhance the participant's performance in navigating routes of varying difficulty, participants perceived it as more useful in helping them to decide where to go, than the system without the ECA.

Then, the video clips used to simulate the routes in the castle were overall perceived well by the participants. However, participants expressed the desire for higher quality video clips and that the video clips should be relevant to the background images used by the systems.

The studies above manipulated the impact of the presence of the ECA on the tasks of information extraction and navigation. They were not designed to take into consideration how the individual ECA attributes impact the user's subject experience and task performance. The following chapter outlines three formal user studies that address this shortcoming and provide further insights on this issue.

Chapter 8 ECA Attributes Studies

This chapter discusses experiments four, five and six of this research work, which aim to provide some insights on the second question of this research, in which aspects/attributes of a multimodal embodied conversational agent influence which aspects of the usability and accessibility of a mobile tour guide system plus the quality of the user experience with such systems. The fourth experiment, examined the impact of two approaches for building natural language question & answering (Q&A) systems (an important attribute of ECA-based systems) and the style of a Q&A session, on the quality of the users' subjective experiences and on their retention performance. The fifth experiment examined the problem of ECA competence and its impact on information retention. The sixth and final experiment, introduces a novel method for evaluating the accessibility of ECA-based information presentation systems. It evaluates the effects of different ECA attention-grabbing strategies, on the retention of information from presentations about the castle of Monemvasia. All experiments (except experiment six, see below) were conducted in simulated mobile conditions, using high-resolution panoramic images.

I first provide an overview of each experiment and the hypotheses to be tested. These hypotheses were generated based on earlier experiments and on short exploratory studies that I conducted prior each experiment, with a limited number of users. Then, I provide an outline of the design of each experiment. Finally, I discuss the analyses that I performed on the collected data and the conclusions that I reached based on the generated evidence.

EXPERIMENT 4

8.1 Overview

The goal of experiment four was to evaluate the impact of two approaches (i.e., script-based vs. parsing) for building natural language question & answering (Q&A) systems, and the style of a Q&A session on the participants' subjective impressions of the answers. In addition, I sought to examine the impact of these variables on the ability of participants to effectively retain information from presentations about the

castle of Monemvasia. The experiment was conducted in the laboratory and manipulated the agent's natural language capabilities (script-based vs. parsing), style of Q&A, and order of task (first route then second vs. vice versa). The script-based system is based on a third-party system, called the **V**irtual **P**eople **F**actory (Virtual People Factory, 2013) while the parsing approach on my own algorithm for processing natural language questions (see §6.5 of Chapter 6, for more details on the two approaches I used for processing natural-language questions). To represent the locations of the castle, I integrated the panoramic scenes used in the previous experiments into the interfaces participants encountered (see Figure 8.1) during the testing.

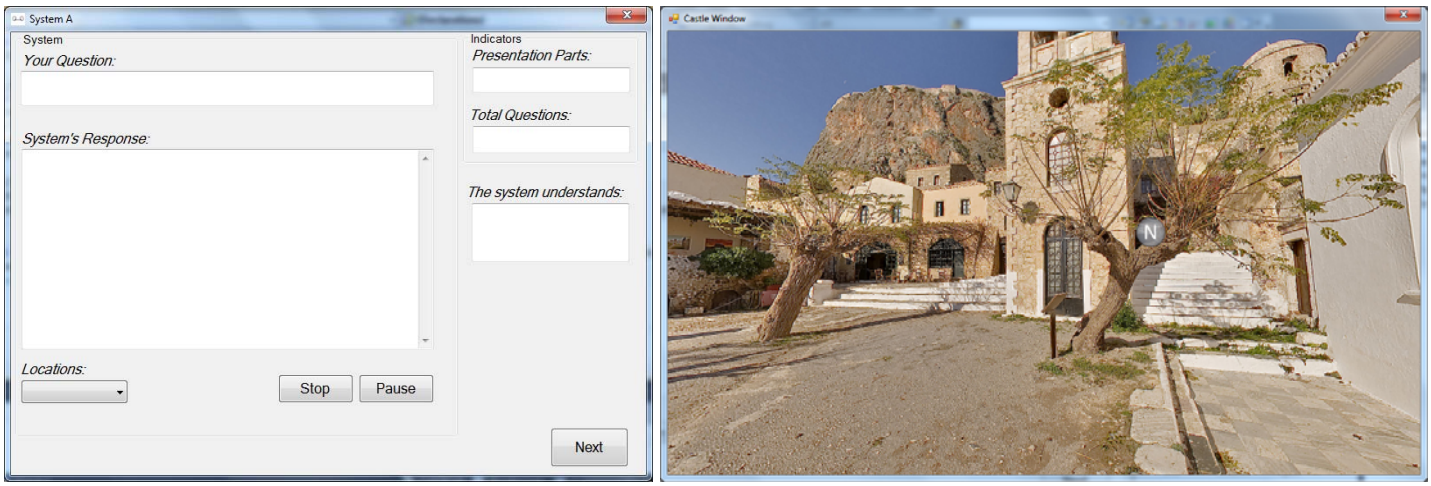


Figure 8.1: One of the two prototype systems with the panoramic window

Based on a short pilot experiment with three users, I formulated the following hypotheses to examine in this study:

User's experience:

H11: Users rate the answers returned by the parsing system better than the answers returned by the script-based system. The parsing algorithm was designed to provide more relevant answers than the scripts, thus resulting in an improved user experience.

H12: If the system does not understand the question, asking the user to rephrase a large number of times benefits retention performance, but not the overall user's

experience. A moderate number (1-2) of questions is more likely to benefit both user performance and the user's experience.

H13: Providing a random answer is better than providing no answer at all. Although a random answer may not be the right answer, it may contain fragments of the information the user is seeking. This should in turn result in an increased user satisfaction as long as this process does not take too much time.

H14: Forcing participants to ask a specific number of questions per location leads to a better retention performance because it forces users to review the provided information multiple times to come up with the questions instead of when users are allowed to ask as many questions as they like per location. However, this effect (forcing participants to ask a specific number of questions per location) impacts negatively the overall user's experience with the system.

Question processing methods:

H15: The deep syntactic processing approach does not, overall, outperform a script-based approach, but it is better for processing more complex questions.

H16: There is no overall performance difference between the script and the shallow parsing approach. However, I expect to find an overall performance difference between the shallow parsing, and the deep syntactic processing approach.

8.2 Experimental Design

In this section I present the design of experiment four. First, I provide an overview of the participants and the software/hardware equipment they used. Then, I present the comparison that I performed between two methods for natural language processing and determine the most robust one. Finally, I present the task participants had to complete and the conditions under which they completed it.

8.2.1 Participants

My approach was entirely user-driven. I initially asked a group of three users to test a preliminary prototype and tell me their requirements. Based on their responses, I refined the original prototype and evaluated it with a group of twelve new participants (see Tables 8.1 and E.4.1 in Appendix E for the participant details). All participants were undergraduate students of Middlesex University who participated for course credit and were randomly assigned to conditions. None of the participants was a local resident of Monemvasia or had visited the castle of Monemvasia before. The participants had a variety of computer science majors and computer experience backgrounds and they were all native speakers of English.

Order of Task	Participants	Age (AVG)	Std. Deviation	Gender (M/F)
First Route vs. Second Route	6 participants	23.1	12.4	6 Males
Second Route vs. First Route	6 participants	24.3	2.3	6 Males

Table 8.1: Table of participants in experiment four

8.2.2 Software and Equipment

For this experiment, I designed two simplistic interfaces: “System A” and “System B”. Each system provided participants with cultural content covering popular attractions on two routes in the castle of Monemvasia and allowed them to ask questions using plain English after each presentation was complete. Each route included three locations to visit in turn (labelled Locations A-C and Location D-F). The systems utilized either the script-based or a parsing-based approach to process natural language questions. An expert human-guide wrote the presentations and crafted the initial conversation corpus using the Virtual People Factory authoring tool (Virtual People Factory, 2013). The interface of both systems is simple enough to use without any previous training and it is divided into the following sections:

- The System section

This section features an input field for typing a question, an output field for displaying the system's response, a drop-down menu for defining the location the user is visiting, and two buttons for controlling the speech output of the system.

- The Indicators section

This section provides information about the total number of questions asked, the database question the system matched the input question to, and the part of the presentation where the user is currently listening.

A simple key combination activates the "Castle Window" that displays an interactive panoramic representation of each location participants had to visit. In case of an unknown input, i.e., the participant asked a question that the system failed to match with the database, the system requested the participant to rephrase the question. If the participant failed to rephrase the question in a way the system could understand a specific number of times (different for each location), the system returned a random answer from the database. This was done to investigate the impact of varied number of times participants had to rephrase a question on their retention performance and experience with the prototypes (see hypothesis H12).

8.2.2.1 Algorithmic Comparison

The algorithm used in the parsing-approach has two layers (a shallow parsing and deep syntactic processing layer) (see §6.5 of Chapter 6), to map the user's input to a proper response in the database (see Figures B.1 and B.2 in Appendix B for code snippets). Although parsing is more precise than the script-based approach, it needs much processing power. As it is not always possible for mobile devices to have a stable internet connection for processing to take place in the "cloud", some of the processing should be conducted locally. For this reason, I sought to compare scripts with shallow parsing (scripts vs. shallow parsing) to get some insights into the robustness of each method. Furthermore, I compared each of these methods with the

deep syntactic processing approach in order to investigate further what is lost when precision is sacrificed for robustness.

Using the end-user logs from both systems, I extracted 60 questions which the systems failed to answer and asked an expert to create their responses. These new sets of stimuli-responses were used to augment the existing corpus using the VPF tool. There is evidence in literature (Rossen *et al*. 2009) that the use of both end-users and domain-specific experts in the process of conversational modelling provides a more comprehensive coverage of the conversational space than when the model is crafted by a developer alone. Therefore, the conversational corpus used by both systems should be sufficient enough to enable a more effective comparison of the methods used for processing natural language questions. The methods that I sought to compare were the following:

Scripts vs. shallow parsing
Scripts vs. deep syntactic processing
Shallow parsing vs. deep syntactic processing

In all conditions, a single user asked each system 60 random questions that covered the four locations in the castle for which the system was providing information and marked each system response using the following scale:

- Each correct answer received 20 points,
- Each relevant answer 10 points,
- Each irrelevant answer 5 points
- No points were given when the system returned a random answer (or no answer at all).

The total score (expressed as the percentage of the points given for achieving a perfect score) achieved gave the overall performance of each method. Table 8.2 shows the results:

Comparison	Performance (out of 100%)
Scripts vs. shallow parsing	59% / 57%
Scripts vs. deep syntactic processing	59% / 57%
Shallow parsing vs. deep syntactic processing	57% / 40%

Table 8.2: Algorithmic performance between the conditions

There was a variation of performance for the deep syntactic processing method, across the content presented about the locations of the two routes (40% vs. 25%) (see Table 8.3 and Table E.4.3 and Table E.4.4 in Appendix E). This effect was not observed in any other condition. This is clearly due to the unknown predicates used in the questions asked in the attractions of the second route. The predicate matching heuristic of the deep syntactic processing layer fails if the database does not contain the relevant predicates.

	Script Approach	Shallow parsing	Deep Syntactic Processing
Locations A - C	58%	56%	40%
Locations D - F	59%	57%	25%

Table 8.3: Algorithmic comparisons per type and location of the tour

In terms of overall performance, the results validate my original hypotheses (see H15 and H16). Although scripts process questions with “poor” language skills, they are more robust in providing overall better answers than the parsing approaches (i.e., both shallow parsing and deep syntactic processing). The slight difference in performance between scripts and the shallow parsing approach shows that the method can be used for filtering-out input-stimuli pairs that do not match grammatically and, therefore, provide more accurate answers. Furthermore, the results show that the deep syntactic processing layer performed below average. As the syntactic parser (Proxem, 2010) is still not “mature” enough, it gave several failed parses of questions that dropped the overall performance of this layer. Therefore, it is reasonable to assume that once the performance of the parser is improved in future versions, the performance of this layer will be improved as well.

8.2.3 Task

To ensure that the systems would run properly, participants interacted with the systems using the Sony[42] Vaio FZ21Z model. After the experimenter provided a brief explanation about the purpose of the experiment, the participants began the task, which was to uncover information about six locations of the castle of Monemvasia and ask questions after the completion of each presentation. They were asked to perform this task once using System A and again using System B. To make it easier for participants to understand the provided information, each presentation was divided into parts and an interactive panoramic representation of each location was integrated in the systems. Participants could interact with the panoramic while listening to a presentation, thus relating the provided information to the actual locations. Half of the participants in each group were told to ask as many questions as they liked per location as long as the total number was not greater than twelve. The other half was restricted to four questions per location. In case the system failed to process one of the questions, participants were asked to rephrase their question as many times as necessary, until they got an answer. Once the system provided an answer, participants were asked to rate thirteen statements on a 10-point scale (1 = no answer 10 = perfect answer). Examples of these statements are clarity and wording of the answers. After the participants had visited all locations, they were asked to write down what they could remember from the presentations (and answers) about each location in total, and freely comment on their overall experience with the systems.

8.2.4 Conditions

The independent variables in this experiment were:

- The approach used for question processing (Scripts vs. Parsing),
- Style of Q&A (forced vs. free), and
- Order of task (first route then second vs. vice versa)

[42] The laptop's full technical specifications can be found at http://www.laptopsdirect.co.uk/Sony_VAIO_FZ21Z_VGN-FZ21Z/version.asp

As dependent variables I measured:

- Performance (i.e., percentage of propositions recalled from the content), and
- The ratings in the subjective impressions questionnaires.

The variable language processing method was manipulated within-subjects (see Table 8.4), whereas the order of task between-subjects. Participants were randomly assigned to the four experimental conditions: 1) script-based system with the first route (i.e., Locations A-C) vs. parsing-based systems with the second route (i.e., Locations D – F) or 2) script-based system with the second route (i.e., Locations D-F) vs. parsing based system with the first route (Locations A-C).

Participants (n = 12)	**Script-based System**	**Parsing-based System**
1 – 6 Participants	**First Route** Subjective impressions/Free recall test/No. of Questions	**Second Route** Subjective impressions/Free recall test/No. of Questions
7 – 12 Participants	**Second Route** Subjective impressions/Free recall test/No. of Questions	**First Route** Subjective impressions/Free recall test/No. of Questions

Table 8.4: The experimental design of experiment four

8.3 Measures and Methods

The only objective variable that was used in this experiment was the accuracy of the answers to the free recall test. The subjective measures were the responses to the items of the questionnaire. The questionnaire items used a 10 point scale (1= no answer 10= a perfect answer) to measure the subjective impression of the participants of the answers provided by the systems. My choice of 10-point scale was consistent with that done in other similar[43] studies (Stevens *et al.* 2006). The

[43] The natural language technology used in the Steven *et al*. (Stevens *et al.* 2006) study is the same technology used in the script-based system in this study.

questionnaire addressed several dimensions of the subjective impressions of the answers such as clarity, sense, fun, etc. (see Table 8.8 for the full list of items)

8.4 Results and Discussion

The following sections discuss the results of experiment four. First, I discuss the results of performance measures. These are the results relating to the order of task, type of system and amount of question asking as independent variables and retention scores as the dependent variable. Then, I discuss the results of the questionnaire and comments participants made after the completion of the task.

Performance Measures

I measured the total number of concepts recalled from the presentations. As a concept I defined one or more sentences that cover the same topic. For example, the following sentence *"The facing on the main gate, like the moulding on the wall, and the corbeling of a small bartizan located to the upper right of the portal, are all made of porous rock, quarried nearby"*, is one concept that covers the material by which the main gate of the castle was constructed. Each test was scored as a percentage of the correctly reproduced concepts. Table 8.5, shows the overall participants' retention performance.

Order of task	System	Mean	Std. Deviation
First route vs. Second Route	Scripts	15.8	16.4
	Parsing	10.8	10.7
Second Route vs. First Route	Script	6.6	4.8
	Parsing	4.5	3.8

Table 8.5: Mean retention performances

A 2 x 2 x 2 ANOVA taking the order of task, type of system, and amount of question asked as independent variables and retention scores as a the dependent variable, showed a significant interaction of order of task and the style of question-asking ($F(1, 16) = 5.245$; $p < .05$) (see Figure 8.2).

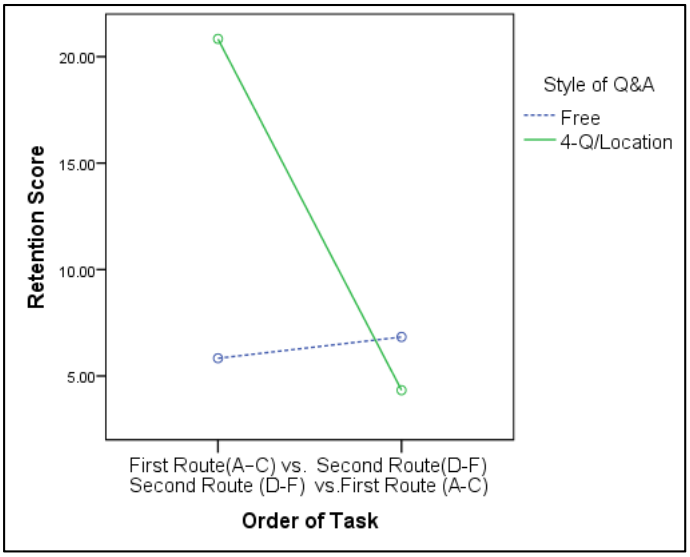


Figure 8.2: The interaction of retention score for order of task and Q&A style

This interaction was further analysed using simple main effects analysis. It showed that the variation of order of task significantly influenced the participants who were forced to ask four questions per location ($F(1, 20) = 10.805$; $p < .01$) but not the participants who were allowed to freely ask questions.

System	Order of task	Questions	Mean	Std. Deviation
System A-Script	First Route (Locations A – C)	Free	7.0	4.5
		4-Q/Location	24.6	20.4
	Second Route (Locations D – F)	Free	8.0	5.2
		4-Q/Location	5.3	5.0
System B - Parsing	Second Route (Locations D – F)	Free	4.6	5.0
		4-Q/Location	17.0	12.2
	First Route (Locations A – C)	Free	5.6	4.5
		4-Q/Location	3.3	3.5
Total	First Route vs. Second Route	Free	5.8	4.4
		4-Q/Location	20.8	15.6
	Second Route vs. First Route	Free	6.8	4.5
		4-Q/Location	4.3	4.0

Table 8.6: Retention performance as a function of Q&A style and order of task

A close inspection of the descriptive statistics (see Table 8.6), revealed that the participants who were forced to ask four questions per location, performed better

overall in the first order (mean A-C/D-F = 20.8), than in the second order (mean D-F/A-C = 4.3). This effect is independent of the type of system used (parsing or scripts). The participants who used the script-based system performed better in the first route (mean A – C = 24.6) than in the second route (mean D – F = 5.3). The participants who used the parsing-based system performed vice versa (mean A – C = 3.3 vs. mean D – F = 17). This finding suggests a correlation between the content participants experienced in each route and the type of question-processing that was used. The content that participants experienced in the second route was domain-specific (i.e., about churches), while in the first segment it was open-ended (i.e., a variation of attractions). Therefore, parsing is a better approach for processing more domain-oriented questions than scripts, while scripts are a better approach for processing more open-ended questions than parsing.

Although there is no significant effect of the system type on the retention scores, it is clear from the table above that participants performed on average better with the script-based system. As scripts were more accurate (see Table 8.2 for the performance of each algorithm), participants got better answers to their questions than when using the parsing system. For every unknown input the system would ask the participant to rephrase the question. This means that the parsing-based system would ask the participants to rephrase an unknown question, more times than the script-based system. I observed in the lab that this annoyed them and most likely distracted them from the information they already had in their minds about the locations. Therefore, the first part of my hypothesis (see H12) is invalid as asking participants to rephrase a question a large number of times does not lead to an enhanced retention performance or improved user experience. Then, based on the participants' comments I argue that the second part of my hypothesis is most likely valid i.e., asking a user to rephrase a question once could lead to improvements in both retention performance and the user's experience. However, apart from the participants' comments I do not have any other evidence to fully support this claim.

In relation to the style of question-asking, the table below (see Table 8.7) shows that the participants who were forced to ask four questions per location performed better overall (mean 4Q/Location = 12.5) than those who were allowed to ask as many questions they liked per location (mean free = 6.3). However, a one-way

ANOVA, testing the difference between the means of both styles failed to reach significant levels ($p > .05$). As it is clear from the descriptive results that participants have the tendency to perform better when they are forced to ask four questions per location, the lack of significance can be attributed to the small number of participants in each group (6 participants / group). In a larger group, it is possible that there would be a statistically significant difference between the participants that used different question styles. Therefore, I argue that there is a strong indication that participants who were forced to ask only a specific number of questions per presentation remembered more information, than the participants who were allowed to freely ask questions.

Style of Q&A	System	Mean	Std. Deviation
Free	Scripts	7.5	4.46
	Parsing	5.1	4.3
4 Q/Location	Script	15	17
	Parsing	10.1	11.01

Table 8.7: Constrained/Free question asking per system

Furthermore, in the lab I observed that those participants got frustrated from having to review the content several times in order to come up with the specific number of questions. Both findings provide grounds that my hypothesis (see H14) could be valid and that forcing participants to ask a specific number of questions enhances retention performance, but not the overall user's experience.

Subjective Assessment

Table 8.8, shows the mean responses for the questionnaire items for the different system and order of task conditions. The questionnaire was highly reliable (Chronbach's a= 0.89). The participants rated all the items of the questionnaire almost similarly. Therefore, my hypothesis (see H11) that the parsing system improves the user's experience by providing more relevant answers is not rejected. Except for "fun" and "accuracy", participants seem to have perceived both methods for processing natural language questions similarly. I performed an ANOVA taking

the participants' ratings for each of the questionnaire items as a dependent variable, and type of system and order of task as independent variables. It showed a statistically significant effect of order of task on the following questionnaire items:

- Item 6 *("Fun")* (F (1, 20) = 4.616; p < .05)
- Item 8 *("Interesting")* (F (1, 20) = 6.943; p < .05)
- Item 11 *("Tiresome")* (F (1, 20) = 12.454; p < .01)

All effects, are clearly because of the variation of content across the order conditions. Participants in the first order, experienced content from the first route (i.e., Locations A – C) with the script system then content from the second route (i.e., Locations D – F) with the parsing system, while participants in the second order experienced the content vice versa.

	(Order 1)		**(Order 2)**	
Measures	**Scripts/Parsing**	**Std. Dev.**	**Script/Parsing**	**Std. Dev.**
Clarity	6.8 / 6.1	2.7/1.9	6.6 / 6.7	1.7/1.0
Wording	6.5 / 6.1	2.5/1.7	6.2 / 6.8	1.7/1.0
Sense	6.3 / 5.8	2.2/1.9	6.0 / 6.8	1.8/0.4
Understandable	6.8 / 6.3	2.2/1.6	6.5 / 7.1	2.0/0.9
Simplicity	6.6 / 6.8	1.0/1.4	6.6 / 6.7	1.5/0.8
Fun	5.8/ 5.0	1.2/1.7	6.5 / 6.7	1.3/1.3
Annoying	2.2 / 2.9	0.9/1.7	2.4 / 2.3	1.3/0.9
Interesting	5.9 / 4.7	1.0/1.7	6.5 / 6.6	1.2/0.5
Intelligent	6.1 / 5.1	1.9/1.7	7.0 / 6.6	1.4/0.6
Stimulating	5.2 / 4.5	1.6/1.8	5.8 / 5.9	0.6/0.6
Tiresome	2.3 / 2.5	0.9/1.1	4.0 / 4.1	1.3/1.3
Unpleasant	2.0 / 2.2	0.7/1.5	2.9 / 2.8	1.2/1.5
Accuracy	6.1 / 5.2	2.4/1.9	6.5 / 7.0	1.9/1.1

Table 8.8: Mean responses to the questionnaire items

Comments:

After participants wrote down what they could remember from the presentations, they were asked to write down openly what they think about the two systems. From the comments participants made, I selected the following and grouped them accordingly.

System A Design (Scripts):

1) Simple and easy to use, with surprisingly accurate answers.
2) Faster than system B

System B Design (Parser):

1) Too slow (it takes up to a minute to load)
2) One participant said that he did not find the answers he was looking for, while another said that this system is more accurate.

The above comments about the two systems are consistent with the patterns of questionnaire scores (see Table 8.8). The script-based system was generally perceived by the participants as faster and more accurate, than the parsing-based system.

General improvements (both systems):

1) If the system cannot answer at least one of the questions, it should take the user back to the same paragraph s/he was reading.
2) Both systems should use easy vocabulary and clearer sentence-structure.
3) When the user enters a question provide suggestions, like Google, to help the user to ask the correct question.
4) If a question cannot be answered, at least the second time, the system should take the user back to the same paragraph s/he was reading.
5) The speed of the text-to-speech (T2S) should be slower.

Participants provided a plethora of suggestions for improvements that can radically enhance the overall user's experience. An improvement of particular importance is the number of times the system should ask the participant to repeat

the question, and the systems' action afterwards. Participants suggested that this should happen just once. The second time, the system should take the participants back to the content it was narrating. This comment provides an indication that returning a random answer (see H13 hypothesis) when the system fails to interpret a question may not be a good idea at all. However, as there is insufficient evidence to support this claim, this issue needs to be investigated further in future experiments.

EXPERIMENT 5

Experiment four addressed the language processing abilities of an ECA without the use of any kind of embodiment. Therefore, I thought that in order to explore the space of ECA features in more detail, an additional experiment was needed where an embodied character is used. In this experiment, I decided to manipulate the competence of the character, that is, the ability to use verbal and non-verbal means to narrate content about the locations of the castle. In particular, I intended to test the following hypothesis:

H17: The behavioural competent ECA enhancement hypothesis: The competent ECA, that is, the ECA that properly uses a range of non-verbal and verbal behaviours to present information augments the participants' ability to retain information from presentations about the locations of the castle. This effect is independent of the type of content (simple vs. complex).

As in the real life guide scenario, I expected that an ECA that uses proper gesturing (i.e., relevant and well-synchronised gestures with the spoken sentences) and suitable pauses between the sentences would be seen by the participants as more than an ECA that displays a minimum set of such behaviours. This should, in turn, result in a better retention of information.

8.5 Overview

In this fifth experiment, I examined how degraded levels of ECA competence affect user performance and, in particular, the participant's ability to effectively retain information from presentations of variable difficult about the attractions of

the castle of Monemvasia. By the term competence, I mean the degree that an ECA uses non-verbal behaviour to augment verbal information. In the context of mobile guides, I view the proper use of non-verbal behaviours as a strong visual indicator of an ECA's competence (see variable competence in §5.3 of Chapter 5). As an auditory indicator of competence, I manipulated the ability of an ECA to properly pause between the sentences of a presentation. As with the previous experiments, experiment five was conducted in the lab under simulated mobile conditions.

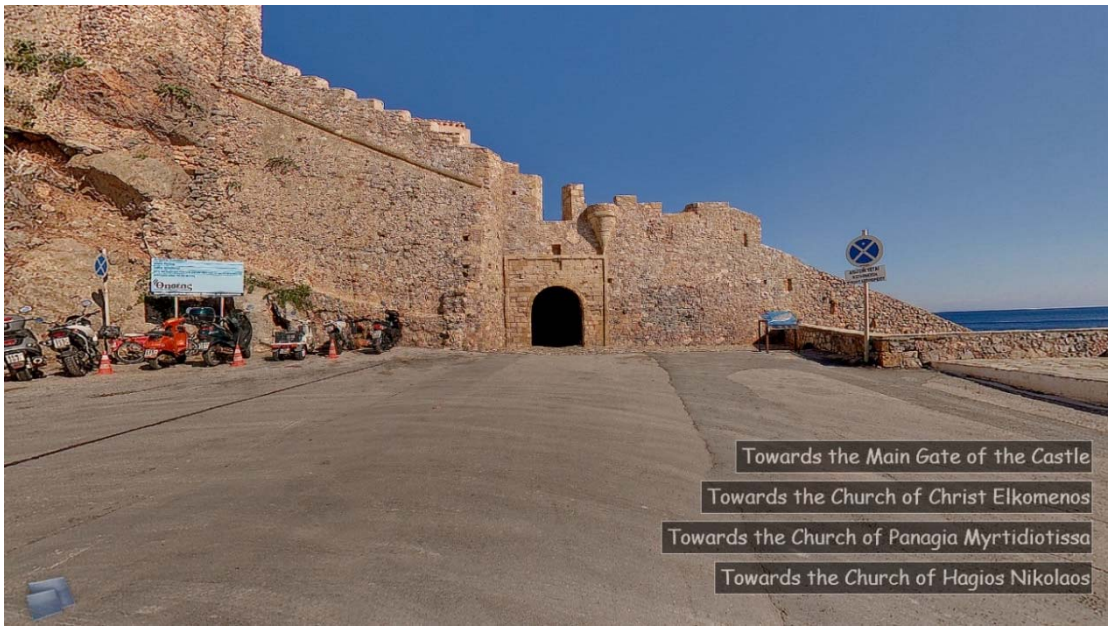


Figure 8.3: A screenshot of the interactive panoramic application

I created a simple panoramic application (see Figure 8.3), representing each location of the castle of Monemvasia that participants had to visit. A simple on-screen menu, allowed participants to navigate from location to location, simply by clicking on the proper menu item. Participants interacted with the panoramas on the same laptop used in the previous experiments, using a wireless mouse.

8.6 Experimental Design

Twelve Middlesex University undergraduates (both males and females) students participated in this study (see Table 8.9 and E.5.1 in Appendix E for the participants' details). The experiment settings were similar to the previous experiments.

Order of systems	Participants	Age (AVG)	Std. Deviation	Gender (M/F)
Fully Competent vs. Low Competent	6 participants	23.8	1.1	5/1
Low Competent vs. Fully Competent	6 participants	25.6	3.0	6/0

Table 8.9: Table of participants in experiment five

I modified the ECA used in the other experiments to create a non-competent ECA with minimal non-verbal behaviour (see right side of Figure 8.4). The non-competent ECA had no pointing gestures and looked away from the user at random intervals (to simulate nervousness). A simplistic control panel provided participants with the ability to repeat a presentation (repeat button) and/or to move to the next presentation (next button). The pauses between the sentences the character spoke were introduced only in the competent ECA. The non-competent ECA spoke continuously without any pauses between the sentences.

Figure 8.4: The fully competent guide (left side) and the low competent guide (right side)

8.6.1 Task

Participants interacted with both systems on the same tablet PC device that was used in the previous experiments. The experimenter provided an explanation of the purpose of the experiment, which was the same for all participants. After that, participants began the task, which was to visit a number of locations in turn, and

uncover information. The presentations referred to different locations of the castle. After participants visited all locations, they had to complete a recall test about the information they heard during the presentations and an object recognition (yes/no) questionnaire on the specific objects they show at each location they visited using the systems.

8.6.2 Conditions

The experiment was a two group, between subjects design. The four conditions varied the competence of ECA (fully competent vs. low competent), the order of systems (fully competent then low competent vs. vice versa), and the type of content (simple vs. complex). Participants were randomly assigned to the four experimental conditions: 1) Fully competent ECA with the simple content vs. low competent ECA with the complex content or 2) Low competent ECA with the simple content vs. fully competent ECA with the technical content.

Participants (n = 12)	**ECA Fully Competent**	**ECA Low Competent**
1 – 6 Participants	**Simple Content** Retention test/Object questionnaire	**Complex Content** Retention test/Object questionnaire
7 – 12 Participants	**Complex Content** Retention test/Object questionnaire	**Simple Content** Retention test/Object questionnaire

Table 8.10: The experimental design of experiment five

8.7 Measures and Methods

The effectiveness of the ECAs in guiding the participants' attention to the objects of the location for which it was providing information about was measured by an object recognition (yes/no) questionnaire. For example, participants were asked to indicate whether they saw a bartizan located above the main gate of the castle. The only objective variable that I measured was the answers to a fill-in-the-blank retention test. The test used the same format as in the previous experiments, i.e., participants had to fill-in a number of words missing from a sentence and to rate the

confidence of their answer on a ten-point scale (1 = completely at random, 5 = not so confident, 10 = totally confident).

8.8 Results and Discussion

This section reports the results of experiment five. These are the results of the object recognition questionnaire, usefulness ratings and the comments participants had after the completion of the task.

Subjective Measures

Table 8.11, shows the total number of the objects participants recognized and those that were missed during the presentations with the two ECAs. A chi-squared test was performed to examine the associations between the type ECA and the type of content with the object recognition responses (yes/no). It showed a highly statically significant association between the type of content and the object recognition responses (x^2 (1, N = 360) = 17.06, p < .001). No other significant associations were found. The participants who experienced the complex content recognized more items in both orders (Total objects = 131) than the participants who experienced the simple content. (Total objects = 93). This effect is independent of the type of ECA used.

Order of systems	**ECA**	**Yes (n = 12)**	**No (n = 12)**
Fully Competent (Simple) vs. Low Competent(Complex)	Fully Competent	47	43
	Low Competent	67	23
Low Competent (Simple) vs. Fully Competent (Complex)	Low Competent	46	44
	Fully Competent	64	26
Total (Y/N)		224	136

Table 8.11: The physical object recognition (Yes/No) results

A plausible explanation of this effect is that when participants experienced the complex content, they focused more on the spoken content ignoring the presence of

the two ECAs. Then, the simplicity of these contents did not require from the participants to focus much on the content. For this reason, they recognised more objects when they experienced the complex content than the simple content.

Usefulness of the systems

After the completion of a session with each of the systems, participants were asked to rate the usefulness of the systems in 1 – 5 scale (with 5 being the highest). The table below (see Table 8.12) shows the mean usefulness ratings. The fully competent ECA was rated overall more useful as a method of delivering presentations than the non-competent ECA. In order to investigate the individual effects of the variables I manipulated, I conducted the following tests: A 2 x 2 ANOVA taking the order of systems, and the type of ECA (Fully vs. Low Competent) as independent variables and usefulness ratings as the dependent variable showed a statistically significant interaction between the order of systems and type of ECA ($F(1, 20) = 20.769$; $p < .001$) (see Figure 8.5). This interaction was further analysed using simple main effect analysis. It showed that the variation of order of systems influenced how participants rated the usefulness of the fully competent ECA ($F(1, 20) = 12.821$; $p < .01$), and the low competent ECA (($F(1, 20) = 8.205$; $p < .05$).

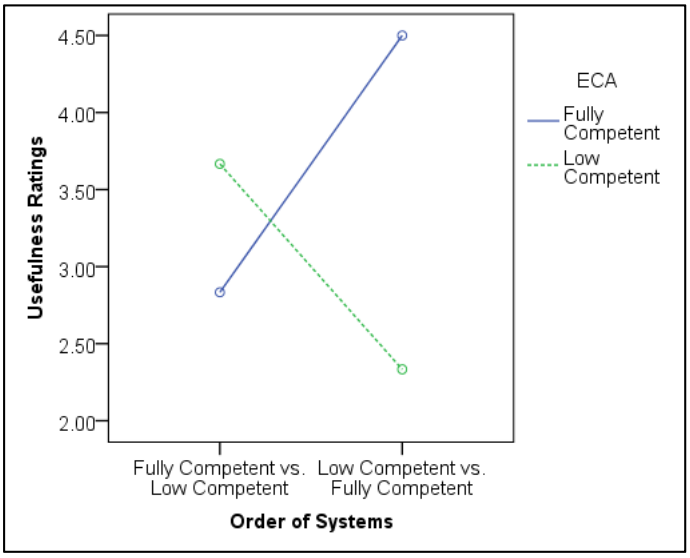


Figure 8.5: The interaction of usefulness for ECA and order of systems

The participants rated the usefulness of the fully competent ECA, higher in the second order (mean Low/Fully = 4.5), than the first order (mean Fully/Low = 2.8) (see Table 8.12). Their ratings for the low competent ECA follow a reverse pattern with participants rating the usefulness of the low competent ECA, lower in the second order (mean Low/Fully = 2.3), than the first order (mean Fully/Low = 3.6).

ECA	Order of systems	Mean	Std. Deviation
Fully Competent	Fully/Low	2.8	0.98
	Low/Fully	4.5	0.54
	Total	3.6	1.15
Low Competent	Fully/Low	3.6	0.81
	Low/Fully	2.3	0.81
	Total	3.0	1.04

Table 8.12: Usefulness ratings as a function of ECA and order of systems

An additional 2 x 2 ANOVA taking type of content (Simple vs. Complex), and type of ECA as independent variables and usefulness rating as the dependent variable showed a statistically significance for content ($F(1, 20) = 20.769$; $p < .001$). The participants rated the usefulness of both ECAs higher when experiencing the complex content (mean Complex = 4.0) than the simple content (mean Simple = 2.5). As participants experienced the complex content across the two orders with ECAs of different type, i.e., with the low competent ECA in the first order and with fully competent ECA in the second order, this explains why their usefulness ratings were affected accordingly. However, as discussed in the comments below, the majority of the participants thought that the fully competent ECA was more effective in disambiguating the spoken information. The gestures used by the ECA made the participants more visually aware of what the ECA was talking about. Therefore, it can be argued that the usefulness of the fully competent ECA was higher than the low competent ECA but this effect is directly related to the complexity of the presented information.

Comments

Participants had to provide an explanation about their usefulness ratings for each of the systems. I divided the participants of both groups, in two groups of equal size. I looked for comments based on frequency and importance in the first group and a corroboration of the comments in the second group (for a full discussion of my approach see §5.2.1.2 of Chapter 5). Below, I discuss my findings grouped into topics. I first discuss the corroborated patterns/comments – those that were corroborated by the participants of the second group. Then, I discuss the uncorroborated patterns/comments – those that were not corroborated by the participants of the second group.

ECA Design:

1) The low competent ECA was better than the fully competent because it did not use many animations. (Uncorroborated comment)
2) The fully competent ECA was better than the low competent guide because it used gestures that made the user more visually aware what she is talking about. (Corroborated pattern by 4 out of 6 participants)

It is obvious from the feedback above that the majority of participants thought the fully competent ECA is more effective in presenting cultural heritage content than the non-competent ECA. Participants explained that the fully competent ECA used gestures that made them more visually involved with the presented content.

3) There is no difference between the fully competent and the low competent guide. (Uncorroborated comment)

The above comment was not corroborated by any of the participants. This shows that the majority of the participants perceived the two ECAs (fully competent vs. low competent) as being different.

Multimodal Content:

1) The content of the low competent system's lack of gestures to signify important information. (Corroborated pattern by 3 out of 6 participants)

The pattern above follows the pattern for the ECA design, and shows that the presence of gesticulation makes it easier for participants to comprehend the narrations of the ECA.

Use of voice:

1) The voice of the fully competent guide was easier to understand (Corroborated pattern by 1 out of 6 participants)
2) The voice of the low competent guide was easier to understand, than the fully competent ECA. (Uncorroborated comment)
3) The voice of the low competent guide was hard to understand because of acoustic problems. (Corroborated pattern by 1 out of 6 participants)

The corroborated patterns above reveal what was rather expected. Participants perceived the content the fully competent ECA narrated with pauses between the sentences as easier to understand than the low competent ECA. Similarly they perceived the content the low competent ECA narrated without pauses between the sentences, as more difficult to understand than the fully competent ECA.

Retention Performance

Participants' retention scores in this experiment are shown in Table 8.13. For this analysis I performed a series of ANOVAs. Participants' scores and confidence ratings were taken as dependent variables and type of ECA (Fully vs. Low), order of systems (Fully/Low vs. Low/Fully) and type of content (simple vs. complex), as independent variables.

Order of systems	ECA	Mean	Std. Deviation
Fully Competent (Simple) / Low Competent (Complex)	Fully Competent	16.6	11.2
	Low Competent	20	15.3
Low Competent (Simple) / Fully Competent (Complex)	Low Competent	15.3	10.5
	Fully Competent	19.5	11.2

Table 8.13: Mean retention scores

I found no significant effects of the variables that I manipulated on either the retention scores or confidence. No significant interactions were found either.

Although no significant effects were found, it can be clearly seen that in the second order participants tended to perform better with the fully competent ECA (mean fully competent = 19.5) than with the low competent ECA (mean low competent = 15.3). Although this result is purely indicative, it is consistent with my behavioural competent ECA enhancement hypothesis (see H17). The competent ECA could augment the participants' ability to retain information and the effect could be independent of the type of content. An explanation for the variation in retention with the fully competent ECA between the two orders (mean Fully (Simple) = 16.6 vs. mean Fully (Complex) = 19.5) can be found in the participants' comments. It is evident that the impact of gestures used by the fully competent ECA was stronger when it was experienced after the low competent ECA that displayed no gestures (see Table E.5.5 in Appendix E). This factor, along with the ability of the fully competent ECA to make the participants visually involved with the content using gestures, is the best explanation for the overall enhanced retention performances with the fully competent ECA in the second order.

EXPERIMENT 6

To date, little research in the ECA community has been conducted using advanced techniques for usability research like eye tracking, let alone, using a technique that combines data from eye tracking, with data from face expression capturing. The human face is one of the strongest indicators of a human's cognitive state and hence how humans perceive stimuli (information, images, etc.). A technique that combines data from face expression recognition and eye-tracking can augment any traditional techniques for accessibility evaluation (e.g., questionnaires, retention tests, etc.). For example, with careful logging one can see which part of the content provided by the ECA system is more confusing, which part requires the users to think more intensively, etc. In addition, eye-tracking data can reveal where the user was looking when a particular expression occurred (e.g., confusion).

In order to validate this new technique for accessibility evaluation of ECA-based information systems, I decided to continue exploring the space of ECA attributes. In particular, I manipulated the strategies an ECA could use to attract attention. This mechanism is an important attribute that presenters must have (not only in the domain of tourist guides) to effectively gain the attention of their audience back, if that has been lost. Although various strategies are possible, I have implemented an ECA that uses either serious or humorous attention-grabbing messages. An automated attention-grabbing mechanism was also implemented in the ECA system used in experiment one (see §7.1 of Chapter 7), but I found it was very difficult to control (some participants would follow what the character was saying, while others not). For this reason, I decided to simulate this feature so that all participants would experience an ECA that uses either one of the two attention-grabbing strategies (serious or humorous). Based on the experience gained from testing the ECA system in experiment one, I had the following hypothesis:

H18: The attention-grabbing enhancement hypothesis: The ECA that uses an attention grabbing mechanism (either the serious or the humorous strategy) enhances the participants' ability to retain information from presentations in the castle. However, it is not known if the participant's retention performance is affected by the serious or humorous content of the attention-grabbing messages.

8.9 Overview

This experiment took place in collaboration with my industrial partner eMarketView[44] (eMarketView, 2013). The company provided the lab, equipment and test subjects and gave me the ability to moderate the sessions remotely. However, in contrast to the other experiments, I did not use an actual mobile device in this study. I designed two ECA systems that run on a desktop computer, simulating a tablet mobile device (see Figure 8.6). I decided not to use the panoramic applications I used in the other experiments to simulate the environment of the Monemvasia castle. This was because of limited resources in my partner's lab, and the difficulty to adjust the eye-tracker to be used on an actual mobile

[44] http://www.emarketview.com/servicios/usabilidad-conversion/eye-tracking

device. The first ECA system employs either the serious or the humorous strategy to attract the attention of the participants to the presentations, while the second had no attention-grabbing mechanisms. The systems provided information about four locations of the castle. Each presentation was designed to evoke at least some content-related emotions such as happiness. I hypothesised that the participant's facial expressions in each presentation would indicate his/her underlying emotional state. Facial expressions were recorded using a camera attached to the computer. Furthermore, I used eye-tracking to identify the section of the interface, where the participant was looking at when the particular expression occurred.

I intended to answer the following questions:

1) Which elements of the interface did the participants look at, for how long and with what face expression(s)?
2) Is there a correlation between the user's behaviour and his/her retention performance?
3) Which elements of the interface did each group of users look at, for how long and with what face expression(s)?
4) Is there a correlation between the above behaviour and the group's retention performance?
5) How does the ECA's attention-grabbing abilities impacts each user's retention performance and potentially his/her behaviour during the interaction? (as this is indicated by the face expressions and eye tracking data)
6) How does the attention-grabbing abilities of the ECA impacts the group's retention performance, and potentially their behaviour during the interaction? (as this is indicated by their facial expression(s) and heat maps)

8.10 Experimental Design

This section reports the design of experiment six. First, I provide an overview of the people that participated in the experiment and their important characteristics (i.e., age and gender). Then I discuss the software/hardware they used, the task they were asked to complete and the conditions under which they completed it.

8.10.1 Participants

The group that participated in this study was composed of thirteen participants (7 women and 6 men) (see Table 8.14 and E.6.1 in Appendix E for the full participant details). I conducted a short pilot study with one of the female participants, to test my approach and calibrate the equipment properly. The remaining twelve participants were assigned equally to the experimental conditions at random. The age range of the group of females was varied to investigate possible age effects.

Order of presentation	Participants	Age	Gender (M/F)
Attention Grabbing (Simple) / Non attention grabbing (Complex)	6 participants	20-30	3/3
Attention Grabbing (Complex) / Non attention grabbing (Simple)	6 participants	20-60+	3/3

Table 8.14: Table of participants in experiment six

In one of the two groups, three women between 35 – 65 years-old were recruited having a normal cognitive ageing with no age associated cognitive decline. All participants were recruited by eMarketView and were paid for their participation. None of the participants had visited the area of the castle before. The participants had a variety of mobile computing and education backgrounds.

8.10.2 Software and Equipment

For this study, I used a modified version of the systems used in experiment five. The attention-grabbing ECA requested the participant's attention in all four presentations, while the non-attention grabbing ECA did not. Although this can be annoying and tiresome for users, especially when the user is paying attention, it was done for two reasons: First, because I wanted to keep the attention-grabbing variable constant across all presentations and second, to investigate the range of emotions that a failure of the attention-grabbing mechanism can elicit to the participants.

Figure 8.6: The Attention grabbing ECA (left side) and the Non-attention grabbing ECA (right side)

Two versions of the attention-grabbing ECA were developed, an ECA that used humorous messages (and the relevant non-verbal behaviours) and an alternative version that used serious messages (and the relevant non-verbal behaviours) to attract the participant's attention. The non-attention grabbing ECA did not use any attention-grabbing mechanisms. A control-panel provided participants access to the presentation of the next location in the tour and to repeat the presentation if desired.

8.10.3 Task

Initially, the experimenter asked the participants to read the experiment brief and to ask any questions they might have. Then they were asked to begin the task, which was to listen to four short presentations in turn about the castle of Monemvasia, once using the attention-grabbing ECA, and the other time using the non-attention grabbing ECA. Half of the participants experienced the attention-grabbing ECA with the humorous messages and then the non-attention grabbing, while the other half the attention-grabbing ECA with the serious strategy and then the non-attention grabbing. In addition, they were informed from the experimenter about the use of the equipment (cameras and eye tracker) and they were asked to participate in a simple calibration task prior the beginning of the task. After participants had listened to the presentations for all four locations, they were provided with a list of randomised keywords and were asked to fill-in a retention test on the information they heard during the presentations. The list of keywords was provided to help participants in recalling Greek names.

8.10.4 Conditions

I measured the type of ECA (attention-grabbing vs. non-attention grabbing) and order of presentation (simple then complex vs. vice versa) as the between-subject variables. The attention-grabbing strategy used by the ECA (humorous vs. serious), the type of content (simple vs. complex) and the participants' gender (females then males vs. vice versa) were manipulated as within-subjects variables (see Table 8.15). Participants were randomly assigned to the eight experimental conditions: 1) Serious AG ECA with the simple content vs. non-AG ECA with the complex content or 2) Humorous AG ECA with the simple content vs. non-AG ECA with the complex content or 3) Serious AG ECA with the complex content vs. non-AG ECA with the simple content or 4) Humorous AG ECA with the complex content vs. non-AG ECA with the simple content.

Participants (n = 12)	**Attention – grabbing ECA (Serious vs. Humorous)**	**Non-Attention grabbing ECA**
1 – 3 Females (20 - 30)	**Simple Content & Serious ECA** Yes/No tests/Retention tests	**Complex Content** Yes/No tests/Retention tests
4 – 6 Males (20 - 30)	**Simple Content & Humorous ECA** Yes/No tests/Retention tests	**Complex Content** Yes/No tests/Retention tests
7 - 9 Males (20 – 38)	**Complex Content & Serious ECA** Yes/No tests/Retention tests	**Simple Content** Yes/No tests/Retention tests
10 – 12 Females (20 – 60+)	**Complex Content & Humorous ECA** Yes/No tests/Retention tests	**Simple Content** Yes/No tests/Retention tests

Table 8.15: The experimental design of experiment six

8.11 Measures and methods

I collected both objective and subjective measures in this experiment. The object recognition (yes/no) questionnaires asked participants to indicate whether they saw specific objects during the presentations by each of the systems. The retention performance of the participants was collected through fill-in-the-blanks tests. The tests followed the format used in the previous experiments, where participants had

to fill-in words missing from sentences, the ECA uttered during the presentations. Below, I discuss the data that I collected from eye-tracking and face-expression recordings. In order to answer the questions posed above, these measures were analysed and the results were correlated with the data collected from the retention tests.

Eye tracker

Gaze trails provided data on which sections of the interface each of the participants cast their eyes, in which order and for how long. This data was collected for each location the user visited using the each of the systems. Furthermore, heat maps provided an amalgamation of where each participant looked at and for how long. I also generated heat-maps for each group of the participants to observe inner-group differences. The "hotter" an area, the more it was noticed by the participants.

Face Recording

As discussed above, I expected that each presentation would evoke a range of emotions from each of the participants. I hypothesized that the facial expressions that participants would display in each presentation would be an indication of their emotional state. As opposed to other studies (Wang and Marcella, 2010) with events carefully controlled to evoke a specific range of emotions (e.g., boredom, surprise, etc.), I did not know what to expect, as the perception of cultural content is a highly subjective experience. However, I expected that participants would display at least some of the basic emotions (e.g., happiness, surprise etc.). For this reason, in a session with either of the systems, I recorded the user's face through a camera attached on the computer.

8.12 Results and Discussion

Object Recognition Questionnaires

Table 8.16, shows the total number of objects participants confirmed (and those that they did not confirm) in the questionnaires under each condition. Overall

participants were able to recognize the objects/artefacts that the ECA included in its narration with a high degree of accuracy. This finding, though, should be considered with caution as the analysis of the gaze trail data produced during the presentations (see *Eye-tracking – Gaze trails*) showed that participants indicated that they saw objects that did not look at on the interface. A chi-squared test was performed to identify any associations between males and females in their ability to recognize objects on the interface. Furthermore, I wanted to see how the type of ECA (attention-grabbing vs. non-attention grabbing), the age of the participants, and the type of content, impacted their ability to recognize objects. I found a significant association between the age of the users and the object recognition responses (x^2 (1, N = 360) = 5.29; $p < .05$). No other significant associations were found.

Order of presentation	**Groups**	**AG (Yes/No) (n = 12)**	**NAG (Yes/No) (n = 12)**
Simple/Complex	Females	25/20	38/7
	Males	35/10	24/21
Complex/Simple	Males	22/23	28/17
	Females	18/27	27/18
Total (Y/N)		100/80	117/63

Table 8.16: The results of the object recognition (Yes/No) questionnaires

The participants of the 35+ female group recognized fewer objects than the participants in all the other groups (see Table 8.16). A close examination of the heat maps of the 35+ group (see Figure 8.11), reveal that participants paid closer attention to the attention-grabbing ECA (and the complex presentation) than the non-attention grabbing ECA (and the simple presentation). The degree of this attention was greater than the participants in all the other groups. Furthermore, participants looked at areas of the interface not relevant to the content of the presentation, which could be interpreted as a sign of "boredom". If the participants during the presentations looked randomly on the screen without actually listening to the ECA's presentation, it is to be expected that they missed a large number of objects to which the ECA was referring.

Difficulty of the Presentations

After the completion of a session with each of the systems, participants were asked to rate the difficulty of the presentations delivered by the ECAs. Table 8.17, shows the mean ratings of the results for the two ECA conditions. A series of 2 x 2 ANOVAs taking ECA type, the order of presentation, attention strategy, and gender as independent variables and the mean difficulty ratings as the dependent variable, showed no statistically significant effects on the difficulty ratings. However, a close inspection of the descriptive statistics reveal that, on average, the female participants of the second group (the 35+) rated the presentations with both ECAs as more difficult (mean rating = 3.3) than the female participants of the first group (under 35) (mean rating = 2.66). Although the difference is not significant, it follows the pattern of the object recognition (yes/no) questionnaires. It is obvious that the older participants may have had a difficult time following the content of the presentation, which clearly explains their higher difficulty ratings. As discussed later in the comments both type of ECAs consisted a source of distraction that deviated older participants from the content presented about the locations of the castle.

ECA (n=12)	Mean	Std. Deviation
Attention-grabbing	3.6	1.08
Non attention-grabbing	3.0	1.00

Table 8.17: Summary of difficulty ratings from E.6.4

Overall it is evident that although participants considered the presentations delivered by both ECAs moderately difficult, they gave better ratings for the non-attention grabbing ECA. An examination of the eye-tracking data, (discussed in the next section) reveals that the attention-grabbing ECA attracted too much attention to itself, which most likely distracted participants from the flow of the presentations. Because of this, participants appear to have perceived the presentations delivered by the attention grabbing ECA as more difficult, than the presentations delivered by the non-attention-grabbing ECA.

Comments

Participants were asked to provide an explanation about how they rated the difficulty of each of the systems. I used the same approach to analyse the interview data as in the previous experiments (for a discussion of the complete approach see §5.2.1.2 of Chapter 5). To enable easier comprehension, I grouped the patterns/comments into the following topics:

- ECA Design,
- Multimodal content design,
- Voice, and finally
- Application design.

Below, under each section I present and discuss both the corroborated and uncorroborated patterns and comments:

ECA Design:

1) The guide that uses the serious interruption strategy is considered rude that prevents users from paying attention to the presentations. (Corroborated pattern by 2 out of 6 participants)
2) The guide that uses the humorous interruption strategy distracts users from the presentations. (Corroborated pattern by 2 out of 6 participants)
3) The guide that uses no attention-grabbing strategies is more effective in attracting the user's attention back to the presentation. (Corroborated pattern by 1 out of 6 participants)
4) The body language of the guide distracts users from focusing on the content (Corroborated pattern by 2 out of 6 participants)

The corroborated patterns about the ECAs follow the overall difficulty ratings, i.e., that the presentations with the attention-grabbing ECA, are more difficult than those with the non-attention grabbing ECA. However, the patterns also reveal that participants did not consider either of the two strategies (humorous or serious) effective in attracting their attention. Curiously, despite the non-attention grabbing

ECA using no attention-grabbing mechanisms, participants considered it more effective in attracting their attention back to the presentations. Most likely, the repeated requests by the ECA for attention made participants annoyed, which eventually led them to lose focus on the presentations. In a fully multimodal system, where the ECA would react based on the real attentiveness of the participants, there would most likely be a clear preference for either of the two strategies. However, even in that case, there is still the issue of the maximum number of times the ECA should request the participants' attention even when the participant is not paying any attention. This experiment provides evidence that requesting attention too many times annoys participants and distracts them from the main flow of the presentations. Based on the data from facial expression analysis (see next section), I suggest a moderate number of times (2-3 times). After that the system, should rely on alternative strategies to regain attention. For example, the ECA could dynamically adapt the presentation flow to closely match the interests of each of the participants. Finally, although the body language used by both ECAs was not excessive, it was considered distracting by the participants. The body language was well synchronized with the speech, but the lips were in some cases out of synchronization. That was a technical problem impossible to solve in the current implementation of the systems, as the avatar engine I used is closed-sourced.

1) The guide that uses the serious interruption strategy is considered rude, but effective in attracting the participants' attention back to the presentations. (Uncorroborated comment)
2) The guide takes a large portion of the screen, which distracts participants from paying attention to the background images. This makes it difficult to remember a narration about a location. (Uncorroborated comment)

The first uncorroborated comment reveals an interesting possibility. The ECA that uses the serious attention strategy could be more effective in attracting the participant's attention back to the presentations than the ECA that uses the humorous strategy. Although the comment was not corroborated, I argue that it is because of the small size of the participant's group. In a larger group, the ratio of participants that would have found the serious strategy effective would have probably been bigger. The last uncorroborated comment reflects what a participant

from the 35+ female group thought. For those participants, as revealed by the relevant heat maps (see the following sections of this chapter), both ECA types were a major source of distraction. It is most likely that ECAs are not a good method for presenting information to older participants, as their use results in increased cognitive workload. However, this issue needs further investigation in future experiments.

Multimodal Content Design:

1) The presentations included a lot of names and historical facts that made it difficult for participants to retain any information. (Corroborated pattern by 2 out of 6 participants)
2) There were too many presentations in a short period of time to retain any information. (Corroborated pattern by 2 out of 6 participants)
3) The Greek names sound similar to users with no relation to the culture of Greece. This in turn, makes it easy to lose the focus of the presentation and forget the information provided about the locations. (Corroborated pattern by 3 out of 6 participants)
4) Once participants have familiarized themselves with the names and the terminology they can acquire more information from the presentations. (Corroborated pattern by 6 out of 6 participants)

In the design of the multimodal content, all patterns were corroborated. In the above list, I can distinguish the pattern about the terminology used in the presentations, and the pattern about the number of presentations. With regards to the first, participants commented that the Greek names/terms, sounded like “nonsense” words that distracted them from the main flow of the presentations. Greek participants did not experience any such difficulty with the content of the presentations. As these terms are culturally-specific, a possible solution to make them universally accessible is to enable the applications for “multimodal associations”. A visual association, i.e., the ability to associate terms/names with images, is a well-known technique for increasing memory performance (Smith and Laurence 2011). Therefore, it is natural to assume that a multimodal association, i.e., the relation of names and terms with more than one modality of communication (e.g., images, speech, gestures, etc.) should result in a superior memory

performance. A possible user scenario is as follows: once the ECA completes its presentation, a list of names/terms used in the presentation could appear. The user could drag and drop the desired name(s) on the ECA that would reply with relevant and appropriate multimodal content (e.g., images, gestures, text, etc.). Finally, I believe that the pattern about the number of presentations is related to the use of terminology that made the participants tired and the presentations to seem long and not to the number of presentations per se.

Voice:

1) The voice of the guide is difficult to understand (e.g., because of interference or speed problems). Use the voice of a real person. (Corroborated comment by 4 out of 6 participants)

Although the text to speech engine (TTS) used in these systems produces one of the most natural speech in the market, participants found it difficult to understand. The use of a real-person's voice could impact significantly the way participants rate the difficulty of the presentations and the information they recall from them.

Application Design:

1) Put the names of the locations in the background pictures to enable easier memorization of the location's name. (Uncorroborated comment)

This last comment shows an interesting method for enabling easier memorization of the location names. Though the pattern is uncorroborated, I will seriously consider it in future versions of the system.

Retention Performance

Table 8.18, shows the retention performance of the participants according to type of ECA and order of presenation. The overall means show that participants, performed better with the non-attention grabbing ECA than with the attention-grabbing one. However, the effect of the manipulation of the various variables (e.g., age, gender, etc.) is not clear in the overall means. Hence, I conducted a series of 2 x 2 ANOVAS taking as independent variables:

- The type of ECA
- The participant's gender and age
- The order of presentation and
- The type of content

And as dependent variables:

- The retention scores and
- Confidence

	AG		NAG	
Order of presentation	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
AG(Simple)/NAG(Complex)	17.6	12.7	20.5	12.8
AG(Complex)/NAG(Simple)	13.0	13.9	23.8	11.9

Table 8.18: Mean retention performances

The tests showed significant effects for both retention scores and confidence. With regards to retention scores, I found a significant interaction of the type of ECA and the participant's gender ($F(1, 20) = 5.845$; $p < .05$) (see Figure 8.7).

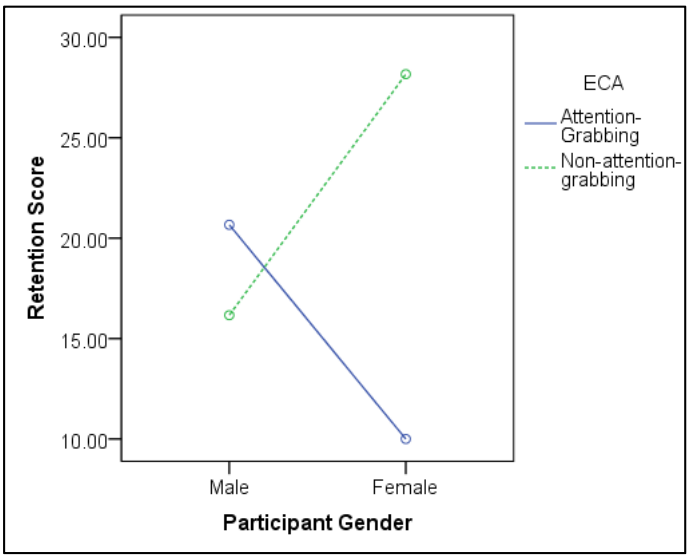


Figure 8.7: The interaction of retention score for ECA and gender

The interaction was further analysed using simple main effects analysis. It showed that the variation of ECA influenced the retention performance of the female participants ($F(1, 20) = 7.509$; $p < .05$) but not the retention performance of the male participants.

A closer investigation of the descriptive statistics (see Table 8.19) reveals that the female participants scored significantly higher when they experienced the presentations with the non-attention-grabbing ECA (mean score = 28.1) than with the attention-grabbing ECA (mean score = 10). The male participants had the exact opposite results. They scored better with the attention-grabbing ECA (mean score = 20.6), than with the non-attention grabbing ECA (mean score = 16.1). The finding for the male participants, is in line with my hypothesis that the attention-grabbing ECA enhances retention performance (see H18), but not for the female participants. An explanation can be found in the facial expression data (see Figure 8.8 and Figure 8.9). In the video files, I observed that the female participants were annoyed by the repeated requests by the ECA for attention.

ECA	Gender	Mean	Std. Deviation
Attention-Grabbing	Male	20.6	14.71
	Female	10.0	9.14
Non-attention grabbing	Male	16.1	8.68
	Female	28.1	12.31

Table 8.19: Retention performance as a function of ECA and gender

This feeling most likely led them to lose focus on the content of the presentation, which in turn resulted in lower retention performances with the attention-grabbing ECA. I did not notice any signs of irritation at the male participants when the attention-grabbing messages occurred. It is obvious that for them, the attention-grabbing messages were more effective in attracting their attention to the presentations. A 2 x 2 ANOVA, taking gender and attention-strategy as independent variables and retention scores as dependent showed no significant effects of attention-strategy on gender. This shows that both attention-grabbing strategies (humorous vs. serious) were effective in attracting the male's participants' attention back to the presentations.

With regards to the possible effects on confidence, I found several significant effects:

- The type of ECA ($F(1, 20) = 17.440$; $p < .01$)
- The order of presentation ($F(1, 20) = 11.480$; $p < .01$)
- An interaction between order of presentation, and type of ECA ($F(1, 20) = 11.267$; $p < .01$)
- An interaction between the type of content and the order of presentation ($F(1, 20) = 17.440$; $p < .001$) and finally,
- An interaction between the type of ECA and the type of content ($F(1, 20) = 17.440$; $p < .01$).

All participants rated the confidence of their answers in the retention tests lower with the attention grabbing ECA (mean confidence = 3.8) than with the non-attention grabbing ECA (mean confidence = 14.9). This shows that participants were actually unsure about whether the answers they provided were the right ones. The low confidence of the answers for the presentations with the attention-grabbing ECA, can explain the low retention scores of the female participants. However, it does not explain the high-retention scores of the male participants. The confidence of their answers with the attention grabbing ECA is too low (mean confidence = 4.83) for the results they achieved.

The remaining significant effects were further analysed using simple main effects analysis. I found the following:

- The variation of the order of presentation influenced the participants who used the non-attention grabbing ECA ($F(1, 20) = 22.747$; $p < .001$) but not the attention grabbing ECA.
- The variation of order of presentation influenced the participants who experienced the complex content ($F(1, 20) = 28.610$; $p < .001$) but not the simple content.

- The variation of content influenced the participants who used the non-attention grabbing ECA ($F(1, 20) = 22.747$; $p < .001$) but not the attention grabbing ECA.

The significant effects above can be attributed to the group of the 35+ participants. The participants of that group scored the confidence in their answers to be very low (mean confidence = 3.66), the lowest from the participants of all other groups. This way, the overall confidence score of all groups in the second order dropped significantly creating the significant effects mentioned above.

Face Recording

A camera attached to the desktop computer recorded the participant's face from a straight angle. I analysed the video files from each presentation separately using a mixture of manual and automated approach. In particular, I used a map of emotions and facial expressions (Joumana 2011) to guide my efforts in observing the relevant facial expressions in the video files. In addition, an automated face detection tool (Fraunhofer Institute 2010) helped this analysis considerably. I observed that the female young participants displayed more facial expressions than the male ones and the older female participants. Figure 8.8 shows the facial expressions of a female participant from the younger group using the ECA with the serious interruption strategy for presentations one to four.

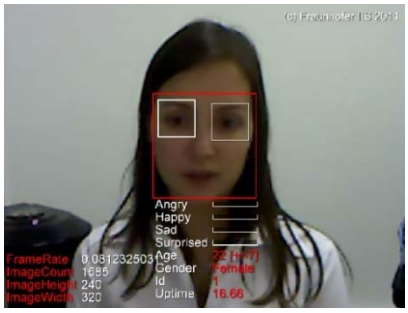
The pictures at the top row show the facial expressions in presentation one. The first picture of each row shows the face expressions of the participant during the presentations. The rest of the pictures of each row show the facial expressions of the participant during the attention-grabbing messages. During the presentation, I observed facial expressions with lips closed and/or slightly up and eyebrows natural. I interpreted these facial expressions as an indication of comfort (relaxed or natural). During the interruption message, I observed face movements that correspond to surprise and happiness. From the comments, the participant reported that she felt at the beginning of the attention-grabbing message that she did not know why the ECA was "yelling" at her, but then she laughed as she was not really paying attention to the presentation.

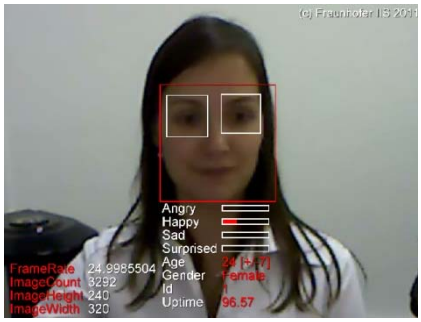
In the second presentation, the participant during the presentation had the same neutral/blank facial expression. However after the attention-grabbing message, the participant changed her facial expression from neutral/blank to an expression suggested increased engagement, with her head turned slightly left, and her eyes open wider compared to before the ECA requested her attention. During the attention-grabbing message, I observed surprise and anger suggesting that the participant was paying attention to the presentation, and she became annoyed by the ECA's reaction. In this presentation, the participant also smiled perhaps to hide the anger or as a self-directed amusement.

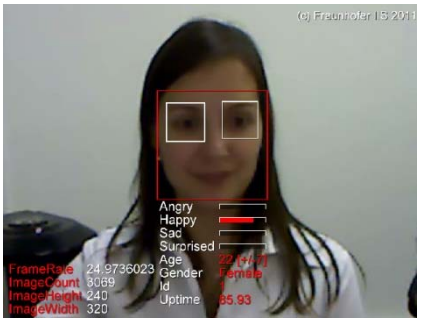
In the third presentation, the participant was engaged in the presentation from the beginning. As in the previous presentation, she was surprised when the ECA requested her attention, smiled for a few seconds, and eventually became angry. It is obvious that she was paying attention to the presentation and that she found the ECA's reaction unnecessary.

In the fourth and last presentation, the participant was again engaged with the presentation from the beginning. Surprisingly, she seemed to find the ECA's attention grabbing message highly amusing. In fact, as it can be seen from the sequence of images below, she displayed one of her most intense facial expressions in all four presentations. A possible explanation is that since ECA requested the participant's attention in all previous presentations, it would most likely ask it again. The feeling of being "watched" constantly most likely created discomfort to the participant, which was masked by a big smile. Also, this feeling most likely distracted the participant from paying attention to the content of the presentation. The participant reported that it was easier to pay attention to the presentations of the non-attention-grabbing ECA than the presentations of the attention-grabbing ECA.

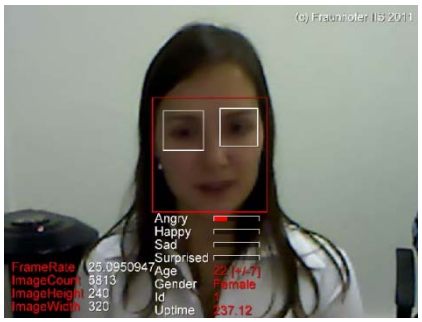
Presentation One

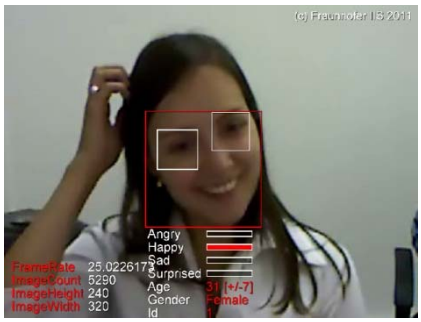


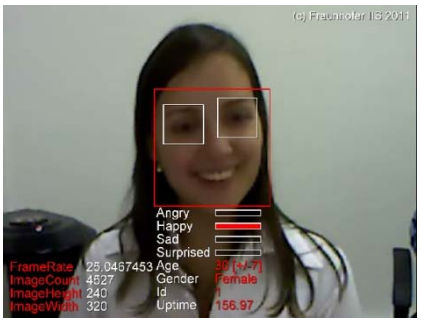


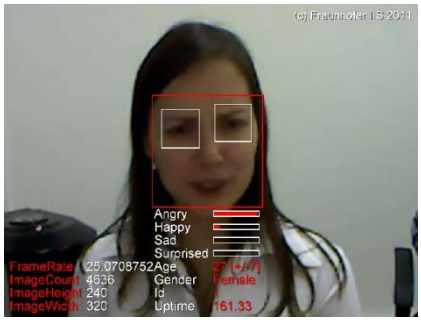


Presentation Two

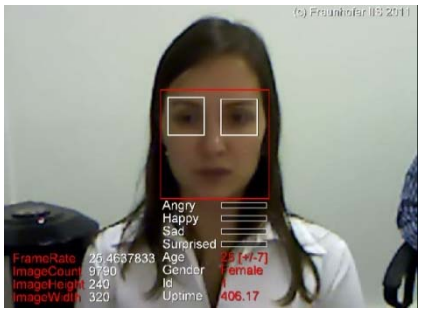


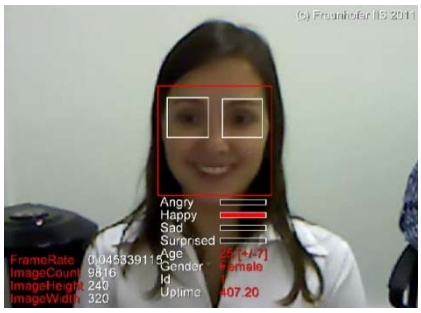


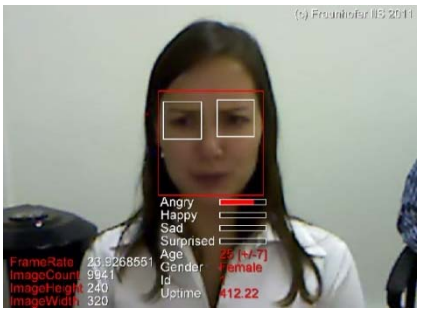




Presentation Three







Presentation Four

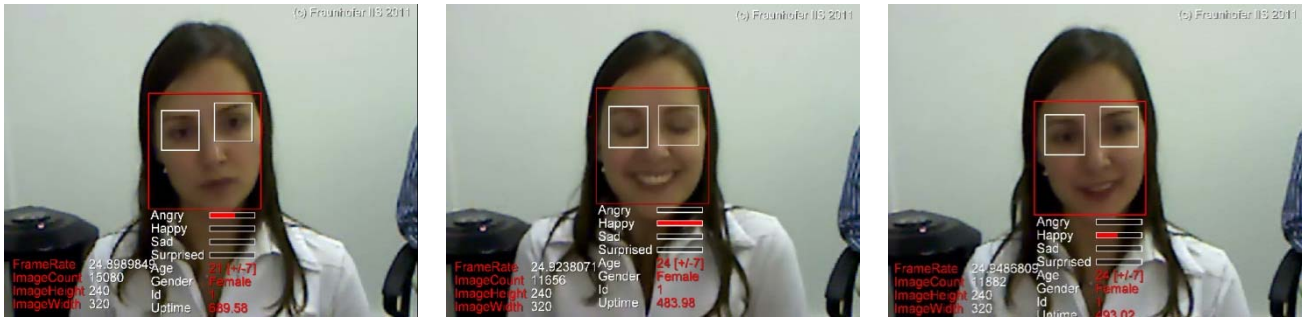


Figure 8.8: Female facial expressions with the attention-grabbing ECA

I did not observe much facial movement in the presentations delivered by the non-attention grabbing ECA. In almost all of the presentations, the participant had a neutral/blank facial expression, followed by a relaxed facial expression. At some points, she also displayed signs of scepticism and happiness, but those can hardly be compared with the intensity and the range of the facial expressions displayed in the presentations with the attention-grabbing ECA.

With regards to the male participants, as was rather expected, it was more difficult for them to show their emotional state using facial expressions. Below I discuss the results of one of the male participants.

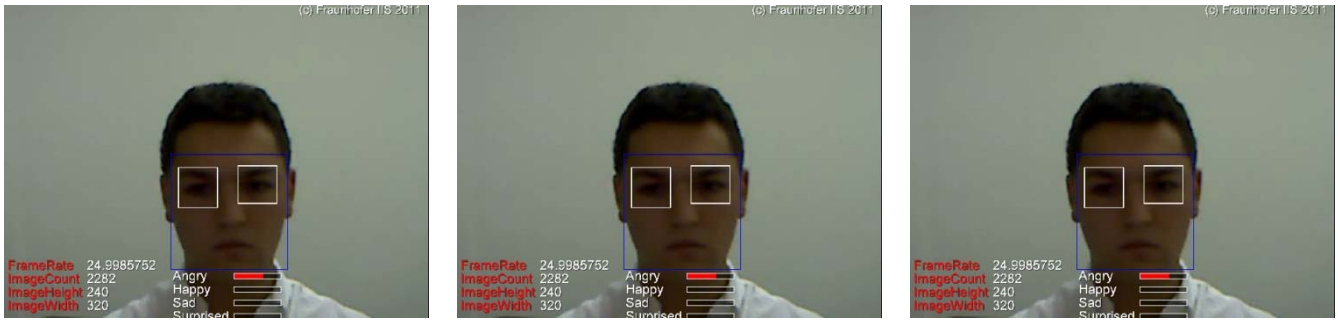
In the first presentation, I did not observe any facial expression change before or after the ECA attention grabbing message. During the presentation, I observed facial movements that correspond to light frown, with eyes fully alert and lips closed and downturned.

In presentation two, the participant slightly smiled after the ECA request for his attention. It is most likely that he found the message of the ECA amusing. However, he almost immediately returned to the previous state of light frown, focusing on the information provided by the system.

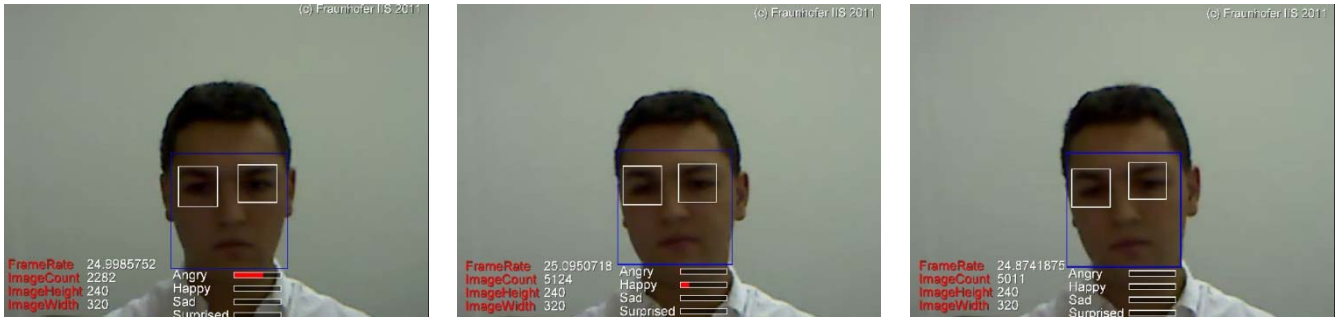
In presentation three, I observed the biggest change in the participant's facial expression. During the attention-grabbing message, the participant displayed his most intense facial expression, a big smile. The most likely explanation is the

sarcastic nature of the ECA interruption message. The participant most likely agreed with the ECA's attention-grabbing message that the content is "boring". The same light-frown facial expressions were observed in presentation four (not shown in Figure 8.9)

Presentation One



Presentation Two



Presentation Three

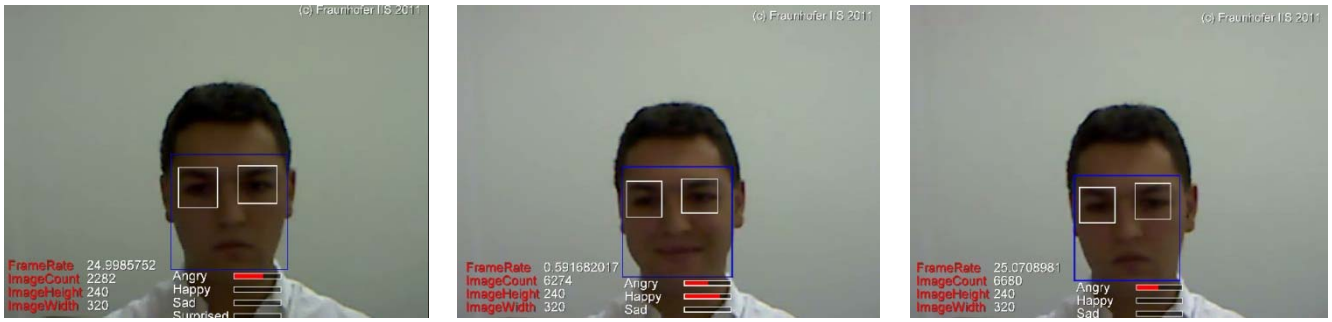


Figure 8.9: Male Facial Expressions with the attention-grabbing ECA

Eye tracking – Heat maps

The eye-tracker produced individual heat maps showing where each of the participants looked at and for how long. The hotter the area, the more it was noticed and looked at by the participants. For comparison purposes, I produced a heat map synthesis of one participant, covering all four presentations (see Figure 8.10) and

attention-grabbing conditions (attention-grabbing vs. non-attention grabbing). The participant was chosen from order one, where there were no differences between the male and female participants. I can clearly see a few clear trends in the distribution of views. In particular, the participant looked at the ECA's face, more than its body features. Then, the background images attracted more attention, than the ECA itself, which shows that the ECA was effective in directing the participants' attention to the objects/artefacts of the background images. However to my surprise, the participant paid more attention to the background images, when she experienced the presentations with the attention-grabbing ECA than with the non-attention-grabbing ECA. This pattern seems to be repeated in all participants of both groups (see Tables E.6.15 and E.6.16 for sample heat maps from the participants of both groups). This effect is independent of the type of content (simple vs. complex) and the attention-strategy used.

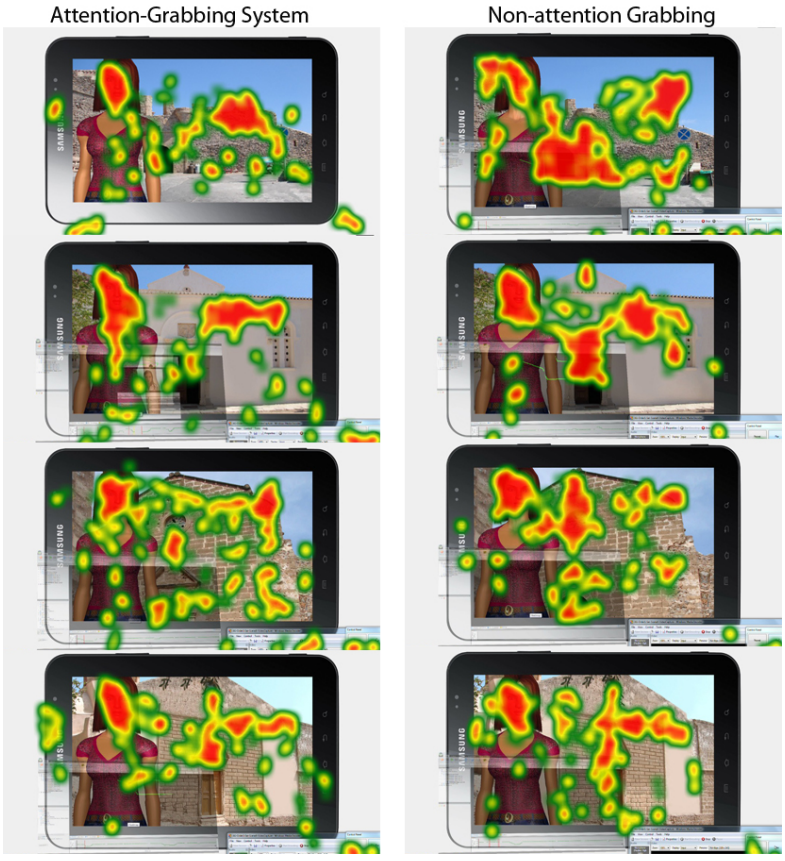


Figure 8.10: Heat maps of one of the participants using both ECA systems

The most plausible explanation is that the attention-grabbing ECA attracted too much attention to it, and hence, distracted participants from paying attention to the

content presented about the location. The constant request of the ECA for attention (regardless of whether the user was paying attention or not), most likely amplified the impact of this effect.

In the second group (see Figure 8.11), I observed a difference in attentiveness between the male and female participants that does not exist in the first group. This variation, though not significant, might be related to the gender of the ECA. However, the examination of the impact of the ECA's gender was outside the scope of this experiment. The most plausible explanation is that this variation is related to the age of the female participants (35+) in this group. It is evident that in all four presentations, the elderly female participants paid too much attention to the presentations with the attention-grabbing ECA, more than the other participants (males and females in both groups). This attention could be interpreted as a sign of increased cognitive activity to follow the presentations, which could explain the negative retention performance of the older female participants (see §8.12).

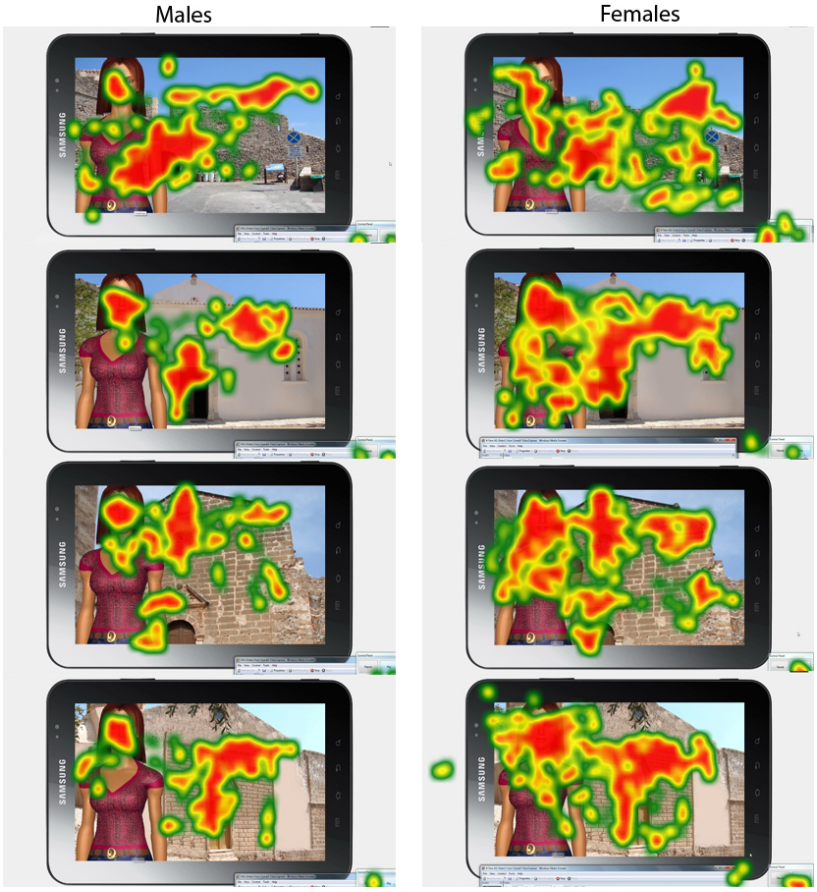


Figure 8.11: Heat maps of two participants (male and female) using the attention-grabbing ECA

The group heat maps show where each group looked the most (both females and males) at each presentation. Figure 8.12, shows a comparison between the attention-grabbing ECA and the non-attention grabbing ECA (in all presentations) for group one. The identified patterns are repeated in group two (see Appendix E.6.17 for a synthesis of the heat-maps of the second group). The combined heat maps confirm some of the patterns identified in the individual heat maps. In particular, there is a general tread of participants to look more at the ECA's face than the body. Also, it is evident that a large portion of the participant's attention was directed by the ECA to specific areas of the background picture.

Figure 8.12: Group heat maps of participants using both ECA systems

However in contrast to the individual heat maps, the combined heat maps show that participants paid more attention to the background pictures, when they experienced the presentations with the attention-grabbing ECA than with the non-attention-grabbing ECA. This could be attributed to the small number of participants in each group (6 people). In a larger sample, the results would have

probably been the same with the individual heat maps, i.e., that the non-attention grabbing ECA attracts the participant's attention to the background images more, than the attention-grabbing ECA. As before, the attention strategy used (humorous vs. serious) had no effect on the attentiveness of the participants.

Eye-tracking – Gaze trails

A gaze trail shows the order in which participants looked each section of the interface and for how long.

Figure 8.13: Gaze trails of one of the participants using both ECA systems

To aid my analysis I divided the interface into two look-zones, the background (including the floating window), and the ECA. As before, I created an illustration to compare the gaze trails of a participant across the attention-grabbing conditions in the four presentations (see Figure 8.13). See Tables E.6.18 and E.6.19 in Appendix E for more samples of individual gaze trails. Again, I can observe some very interesting patterns. First, it can be clearly seen that the attention-grabbing ECA

attracted more attention to itself than the non-attention grabbing ECA. This effect is independent of the attention strategy used (humorous vs. serious), and most likely distracted participants from paying attention to the background images. Participants using the non-attention grabbing ECA, most likely paid more attention to the background images than the ECA itself.

To verify these qualitative findings, I conducted a series of 2 x 2 ANOVAs taking the total number of fixations and time per participant as the dependent variables, and type of ECA and look zone as the independent variables. I found a highly significant effect of ECA on the total time ($F(1,188) = 17.661$; $p < .001$) and number of fixations ($F(1,188) = 33.549$; $p < .001$), and a highly significant effect of the look zone on the total time ($F(1,188) = 159.840$; $p < .001$) and number of fixations ($F(1,188) = 79.994$; $p < .001$). It can be clearly seen from Table E.6.6 (see Appendix E) that participants paid more attention to the system with the attention-grabbing ECA (mean number of fixations = 102.42, and mean time of fixations = 40.18 sec) than to the system with the non-attention grabbing ECA (mean number of fixations = 67.5 and mean time of fixations = 31.30). Furthermore, all participants, regardless of the type of ECA they used, paid more attention to the background images, than the ECA itself. However, the attention-grabbing ECA was more effective in directing the participants' attention to the background (mean number of fixations = 134.2 and mean time of fixations = 54.8), than the non-attention grabbing ECA (mean number of fixations = 89.5, mean time of fixations = 43.3). Hence, my prediction based on the gaze trail images that the attention-grabbing ECA distracted participants from paying attention to the background images is not supported.

The above findings, contradict the attention-grabbing argument against the use of ECA on computer interfaces (e.g., Walker *et al.* 1994). An ECA that uses verbal and non-verbal means to communicate information can divert the user's attention away from itself and towards objects of interest in the interface. However, there is still the issue of how effectively this focus guidance was done and whether participants look at the objects/artefacts pointed by the ECA in the background. To answer this question, I examined the matches between the objects participants said they saw in the questionnaires and those they fixated on the interface. I found that there is a

match between the questionnaires and the relevant gaze fixations for most of the objects/artefacts. For example in Figure 8.14, the participant has cast her gaze upon the "Christian Cross" at the top of the church several times, which shows that she understood what she has seen and explains why she confirmed the object in the questionnaire. However, it is unknown whether the participant successfully connected the object/artefact with the information presented by the ECA. The inclusion of visual questions (e.g., what is the name of this object) in the retention tests could help us provide further insights into this question. Second, there is a mismatch between the questionnaires and the relevant gaze fixations. Some of the participants said they saw an object in the questionnaire, but they did not cast their gaze on the relevant objects/artefacts on the interface.

Figure 8.14: Fixations on an object confirmed in the questionnaires

A likely explanation about this mismatch, is that participants confused these objects/artefacts with other similar objects/artefacts they saw (and cast their gaze upon) on the interface. However, it can be argued that the ECA directed effectively the attention of the majority of participants to the majority of the objects on the interface it was providing information about.

Last, but not least, as the panoramic applications were not included in this experiment, the participants' attention was focused on the interface itself. It is unknown if participants would have been able to find any objects/artefacts if their attention was split between the screens of the systems and the panoramas. Then, if

the participants were in the actual castle of Monemvasia, they would most likely have even bigger difficulties to find the objects/artefacts the ECA was referring to. This is because external environmental factors (e.g., lighting conditions, etc.) would have affected their ability to relate objects/artefacts the ECA is pointing at on the interface of the device with those in the actual environment.

Correlation (Facial Expressions, Gaze trails, retention tests)

I performed a correlation between the facial expressions, gaze trails (number and total time of fixations per section of the interface) and retention tests (see Tables E.6.7 to E.6.14 in Appendix E) to help explain particularly good or bad performances in the retention tests. From both groups of participants, I chose one example of particularly good and bad performance in the retention tests. Then, based on the correlated data I attempt to explain the outcome.

	Group 1 (Simple/Complex)		**Group 2 (Complex/Simple**	
Performance	**AG (S/H)**	**NAG**	**AG (H/S)**	**NAG**
Bad	0%	29%	7%	11%
Good	33%	29%	36%	17%

Table 8.20: Sample retention performances

Table 8.20, shows the retention performances of two participants, and the conditions under which they tested the systems. The retention samples selected for the attention grabbing conditions reflect both attention-grabbing strategies (humorous and serious). Beginning with the bad performances, it can be clearly seen that the selected participants remembered very little when they watched the presentations with the attention-grabbing ECA (humorous or serious). A careful examination of Table 8.21, reveals that the attention-grabbing ECA attracted more attention to itself and the background than the non-attention grabbing ECA (measured in terms of total number and time of fixations). The difference between the two types of ECA was statistically significant only for the total time of fixations ($F(1, 28) = 5.436$; $p < .05$). Furthermore, at each presentation until the interruption

message occurred the participants had either neutral/blank or attentive facial expressions, which shows that they were attentive to the presentations. However, when the ECA requested for attention, I noticed the following: a) the first time the ECA requested attention, participants were either surprised or curious, possibly because they did not expect the ECA to observe their behaviour and b) the initial emotion degraded gradually in every presentation when the ECA asked for attention. In fact, one of the two participants got annoyed at the second presentation when the ECA asked for her attention again. This was most likely because she was already paying attention to the presentation. It is obvious that the repeated interruptions diverted participants from the flow of the presentations, which in turn distracted them from keeping the content in mind.

Bad Performances								
	Background				**ECA**			
ECA	**Number of Fixations**		**Total Time of Fixations**		**Number of Fixations**		**Total Time of Fixations**	
	Mean	**Std. Deviation**	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
AG	125.2	48.6	59.4	14.4	66.8	29.7	29.9	8.9
NAG	102.2	57.2	43.1	25.0	52.5	36.6	19	13.4

Table 8.21: Mean fixations and times of participants with bad performances

Last, but not least, the feeling of being under constant surveillance by the system (the “big-brother” effect) may have made participants nervous, thus diverting them from recalling any information.

With regards to good performances, participants remembered a moderate amount of information from the presentations. A careful examination of the data in Table 8.22 reveals the following about the two participants. The interaction between the type of ECA and the look zone (i.e., background and avatar) was significant for the total number of fixations ($F(1, 28) = 8.007$; $p < .01$). A simple main effect analysis showed that the attention-grabbing ECA attracted significantly more attention to the background ($F(1, 28) = 5.939$; $p < .05$) than the non-attention grabbing ECA. The

mean number of fixations differences for the type of ECAs did not reach statistical significance levels. Until the interruption message of the attention-grabbing ECA, participants had a neutral/blank face, which shows their attentiveness to the content of the presentations. Nonetheless, in contrast to the other two sets of participants, their facial expression during the interruption messages were more constrained (e.g., a slight smile, or even neutral/blank).

Good Performances								
	Background				**ECA**			
ECA	**Number of Fixations**		**Total Time of Fixations**		**Number of Fixations**		**Total Time of Fixations**	
	Mean	**Std. Deviation**	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
AG	132.2	43.1	51.4	14.07	28.4	10.8	21	14.1
NAG	89.3	43.9	42.3	20.0	56	32.3	23	10.0

Table 8.22: Mean fixations and times of participants with good performances

This most likely means that those participants were not distracted by the attention-grabbing messages and they were able to keep their focus on the presentations. In addition, the patterns observed in the participants with the bad performances when the ECA requests for attention (see above) were not observed in these participants. Apart from the slight reactions, the attention-grabbing messages did not have any effect. This provides a reasonable explanation for their good performances in the retention tests.

Based on the above discussion, although with caution, I argue that my proposed method for evaluating the accessibility of ECA-based information presentation systems, works. However, as the method relies on data from multiple sources, it is difficult to produce deliverables in a reasonable time. The task becomes even more difficult when the data have to be captured under mobile conditions. With regards to the software needed, the “Instruments Building” module of the Talos toolkit (see §6.3.6 of Chapter 6), could be enhanced to include all the necessary software needed for the analysis of the video files and eye tracking data. In relation to the

hardware, mobile eye trackers and wearable video cameras can be used to capture the necessary data under mobile conditions.

8.13 Conclusions

The fourth experiment provided evidence that retention of cultural heritage content is related to the accuracy of the method used in the question-answering session with the system. The more robust the method is, the less are the chances for participants to become distracted and forget what they heard in the presentations about each of the locations. In a comparison between three approaches for natural language processing, scripts seem to be overall the more robust approach. However, I found evidence that parsing is better for processing more domain-oriented questions (e.g., the architecture of churches) than scripts. In addition, as the technology for deeper language understanding is steadily improving, parsing approaches may prove to be superior in the future. Then, I found a strong indication that when retention performance is the desired output of the interaction process with a Q&A system, participants should be required to ask a specific number of questions per location. However, this approach is frustrating for users as it forces them to review the content many times in order to come up with the required number of questions. A last important finding has to do with the requests to rephrase a question when the system fails to match it. I found that the repetitive requests annoy participants and affect their retention performance. Therefore, to ensure an optimal Q&A session, the request should be repeated just once, as participants suggested, or the repeat messages should be built in a way to allow users to figure out how to ask the system questions to avoid improper responses.

In the fifth experiment, I found evidence of the importance of the ECA's body language and pauses in speech in presenting cultural heritage content. Although the scores for retention across the two ECAs were not statistically significant, the majority of participants indicated that the gestures used by the ECA and the pauses in the narrations made them more visually involved with the content it was presenting. It can, therefore, be argued that the mere presence of an ECA on an interface is not enough to engage participants with the content. For participants to

consider the ECA as part of the interaction process, it needs to augment the cultural content with relevant and well-synchronised verbal and non-verbal behaviours.

In the sixth and final experiment, I found strong quantitative and qualitative evidence that an ECA should not attract attention to itself more than necessary, to avoid becoming a distraction from the flow of the content. I found proof that an ECA with attention-grabbing capabilities (humorous or serious), can effectively divert the participants' attention focus to relevant content-objects in the background. However, I also found that the attention-grabbing messages had a detrimental effect on the overall retention performance. This affected the female participants more than the male participants. The male participants performed better with the attention-grabbing ECA, while the female participants did so with the non-attention grabbing ECA. Therefore, even when the system can actually react to the user's attention state, a minimum threshold where the ECA can request for the participant's attention must be established. This threshold is difficult to determine as the ECA must avoid becoming tiresome, but also must be effective enough to attract the participants' attention back to the presentation when it has deviated.

Then, I found strong evidence that the use of ECAs on computer interfaces is not a good idea when the system is designed for older participants. The multiple communication channels used by the ECA result into an increased cognitive overload that makes it difficult to readily follow the content of the presentations. I argue that for those participants who are overloaded it is much easier to use more "traditional" channels of communication when designing mobile guide systems, such as text and voice.

Overall the sixth experiment produced substantial evidence to support the claim that my proposed method for evaluating the accessibility of cultural heritage content actually works. The method combines data from face expression analysis, eye-tracking and retention tests to provide a high-quality alternative to the more expensive and more unpleasant method of measuring the user's brain activity (e.g., Simple Usability, 2013)

Chapter 9 Conclusions and Future Work

This final chapter summarizes this thesis work by discussing its contributions to the research community and to knowledge. Following the summary, the chapter concludes by discussing opportunities for future research based on the findings presented in this work.

SUMMARY OF SIGNIFICANT CONTRIBUTIONS

I believe that this thesis provides significant contributions in multiple domains of the ECA research community. In particular:

- **Technical Contributions**

First of all, the Talos authoring toolkit generated from this work (see Chapter 6) allows researchers to design and evaluate ECA prototypes for mobile devices more efficiently. This toolkit is provided with a full information architecture (IA) and documentation that should assist towards its implementation. In addition, I generated a number of design recommendations (based on the evaluation of similar toolkits) that should guide its actual user interface (UI) design. The toolkit's architecture is extendible and provides a number of modules that makes system implementation and evaluation a much easier job, with the possibility of adding more modules as needed. Such a toolkit requires a considerable investment of effort and it is not easily duplicated.

Second, from the Talos modules I partially developed its natural language processing component with the aim to use it in one of my prototypes. The resulting search-and-matching three-tier algorithm provides a more robust and linguistically motivated different option to the open-source tools currently available to the ECA research community. I have plans to make the algorithm open-source, hoping that other researchers will find it interesting and development will continue.

- **Conceptual framework and design recommendations**

The research framework I developed as part of this research work can help facilitate research in the use of ECA's in mobile guide applications. It provides researchers with a common pillar to accumulate compare and integrate results from different studies. Although the framework is domain-oriented, it can be useful to researchers in other relevant domains too (e.g., mobile e-learning). I used this framework to design and implement six empirical studies that generated a substantial number of design recommendations. The recommendations are not by any means definitive and comprehensive, but are aimed at future researchers who intend to take the work reported in this thesis further.

A total of 41 recommendations are presented in detail below. The recommendations are based on quantitative and qualitative evidence generated from my experiments and substantial expert knowledge and experience gathered from the design and development of a number of ECA-based guide applications for mobile devices. They are presented in layman's terms and are divided in the following categories:

ECA- Design

Build a 3D model:

1. **Prefer 3D photorealistic avatars over 2D cartoon-like avatars**
   A 2D character by today's standards is most likely to be considered out-of-date.

2. **For increased avatar realism, use a real-time avatar engine**
   Previously rendered video files of ECAs look cumbersome, and hinder the overall user's experience.

3. **Enable user-modified avatars**
   Enable users to modify the appearance of the ECA to match their individual preferences and needs. For example, the results from my experiments in Greece (see experiment one and experiment two) suggest that users would prefer an ECA with a more Greek-like appearance.

4. **Optimize your character for mobile use**

   Reduce the number of polygons used as much as possible without losing character quality. High-polygon characters take much CPU and graphics resources and hinder overall system performance.

5. **Consider using clothing textures instead of 3D models of clothes**

   Although 3D models of clothes add an increased realism to the character, they add to the overall polygon count and they should be avoided.

6. **Make your textures as realistic as possible**

   Use high-resolution photographs of real clothes, skin, hair, eyes, and teeth to create textures that are lifelike and realistic.

Non-verbal behaviours:

7. **Prefer an ECA with non-verbal cues over an ECA without such cues**

   I found evidence (see experiment five in chapter 8) that body gestures and facial cues (e.g., smile, eye contact) make the user more visually involved with the content than not having these cues at all. This can potentially lead to a better memorization of the content.

8. **Each ECA gesture should match accurately what is being said verbally**

   Participants in all of my experiments suggested that the body language of the ECA should be improved. An ECA that displays asynchronously to the speech gestures distracts users from the content it presents.

9. **Avoid manually creating animations for quick prototyping of a virtual guide system**

   Body language (beat and deictic gestures, etc.) and facial expressions can be created comparatively quickly using new and inexpensive motion-capture devices (e.g., Microsoft Kinect).

10. **Create a minimum of twenty different mouth and lip positions to achieve realistic and natural lip synchronization**

    My experience suggests that this is the minimum number of visemes (visual representation of phonemes), that the character should have for high quality lip synchronization.

11. **Ensure one gesture per sentence of the content**

    If you attempt to synchronise more than one gesture per sentence of the content, the gestures will most likely overlap making the character look unrealistic.

12. **Avoid displaying negative facial expressions**

    Although participants in my experiments did not notice the character's negative facial expressions, it is best to avoid them as they make the character look unrealistic. This is because the current ECA technology cannot offer consistent levels of behavioural fidelity to avoid the uncanny valley effect[45] (MacDorman & Ishiguro 2006). Adding negative facial expressions to an ECA, that looks like a human being, but does not exactly behave like one, will increase this effect and make the character look revulsive and less natural.

13. **Ensure optimal reactions to multimodal input**

    The ECA should respond (e.g., to request the user's attention) to multimodal input with a maximum 1 second delay and without the user's intervention. In any other case, enable users to turn-off the reactions and move on with the narration.

Voice:

14. **Avoid using the same voice tone**

    Vary the way the ECA speaks to add realism to the narration about a location. For example, the ECA could talk faster, slower or stop speaking to attract attention.

[45] The Uncanny valley effect state that when human-like objects that look and act almost, but not perfectly, like human beings cause an unpleasant impression to people.

15. Adopt a moderate rate of speaking

If you are using a text-to-speech engine, make sure you adjust its settings to create a moderate rate of speaking. If the voice still feels fast and unrealistic, introduce a 1s pause between the sentences of the content**.**

16. Prefer the voice of a real-human

The results from my studies suggest that participants prefer the voice of a real human instead of a voice generated by a text-to-speech engine.

Multimodal Content design

17. Have an expert human author to create the cultural heritage content

Do not attempt to create the content based on a guide book. The content should be created by a human author for everyday spoken use.

18. Provide as much visual support as possible for the designer to tag the content

Ideally, this support should be in the form of real-time execution of the tagged content.

19. Prefer historical content, when quick deployment of a virtual guide system is needed

My studies suggest that the majority of the users prefer to explore historical content over other types of content when they visit a historical attraction.

20. Unless you can simplify it, avoid authoring content that requires certain user expertise

The results of experiment one, (discussed in Chapter 7) suggest that an author should avoid creating cultural content that requires certain user expertise (e.g., Architecture).

21. Do not overwhelm users with historical dates

An ECA narrating some historical facts shows credibility, but do not overwhelm users with too many historical dates.

22. Personalize the content when necessary

Include opinions and/or personal experiences to make the content more realistic. For example, *"the construction of the sea wall of the castle really catches your breath, don't you think?"*

23. Author content that is "unexpected" and "spontaneous"

Some examples are: a humorous answer to a question posed by the user or an unexpected reaction to the lack of user attentiveness to a presentation.

24. Author content that is concise and right to the point.

Qualitative evidence from my experiments shows that the content should have short sentences without unnecessary information (e.g., be polite). Keep the navigation instructions to the absolute minimum and make allowances when narrating content about a location.

25. Create content that matches the time-constrains of the visitors

The content should reflect the time-constraints of the visitors. For example, short-stay visitors, could experience content of general interest about the castle. Long-stay visitors, on the other hand, could experience more elaborate content about the castle.

26. Author content of general interest

The results from my first experiment, suggest that participants would like content of general interest (e.g., about local shops, dances, etc.) to be included in the list of the available information scenarios.

Mobile guide application design

27. Avoid using menu-based dialogues as a means of Q&A (Question &Answering) with the ECA

Unless built with an authoring tool, my experience suggests that menu-based dialogues are very time consuming to properly construct, debug and they are not easily extendible. The use of natural language input is highly recommended as an alternative.

28. **Design for touch interaction and for the latest generation of mobile hardware**

    Ensure the user interface elements (e.g., buttons, dialogue menus, etc.) on the application are large enough to avoid the "fat-finger" effect. Furthermore, avoid using hover effects (e.g., to highlight interface items) and right click menus. Finally, design your system layout for both portrait and landscape orientations.

29. **Use text as an additional output modality when presenting technical content**

    I found evidence in one of my studies (see experiment 2 in Chapter 7), that the careful use of text as an additional output modality when presenting technical content can be beneficial for users. I recommend the style of text used in my experiments, that is, auto-scrolling subtitles synchronised with the rest of the content.

30. **Avoid using text when giving navigation instructions**

    I found no evidence (see experiment 3 of Chapter 7) that using text when giving navigation instructions is beneficial for the users. On the contrary, an ECA with a relevant body language was perceived as more useful in helping participants taking the correct navigation decisions.

31. **Use content-enabled objects**

    When the length of the ECA's main narration about a location must be reduced, use content-enabled objects placed on the background of the character. For example, an ECA can give a basic narration about a location, and then "invite" users to explore more by touching the objects they are interested to learn more about.

32. **Enable your mobile guide applications for QR-Code recognition**

    QR-Codes provide a cheap alternative to more expensive and complex solutions such as GPS, for physical location tagging.

33. Give users control over the content and the visibility of the ECA

Give users high-level of control over the content and the visibility of the ECA. The minimum control users expect is a repeat and pause functionality. Then, an option to control the visibility of the ECA would benefit greatly for example, elderly users. I found strong evidence (see experiment 6 in Chapter 8) that ECA-based presentation systems cause increased cognitive workload to elderly participants, making it difficult for them to follow the narrations.

Interaction Design

34. Prefer providing a random answer in Q&A, than no answer at all

The participants' reactions I observed in one of my studies (see experiment 4 in Chapter 8) reveal that participants are distracted by repeated requests to rephrase an unknown to the system question. This impacts the participants' retention of narrated content about the locations of the castle.

35. Do not limit the number of questions per location of a tour

I found a strong indication that when users are forced to ask a specific number of questions per main narration of a tour, their retention performance increases. However, I also observed that their perception of the friendliness of the system decreases. Therefore, unless the desired outcome of the interaction with the mobile guide system is enhanced retention performance, allow participants to ask as many questions as they like per location of a tour.

36. Consider alternative types of interruption messages for attracting attention

Quantitative and qualitative evidence (see experiment 6 in Chapter 8) suggests that both serious and humorous messages worked for attracting attention, but not in a user-friendly way. Hence, a combination of both humorous and serious messages will most likely work more effectively than each of the strategies alone. This is because mixing serious with humorous messages could result in diffusing any feelings of discomfort while attracting an adequate level of attention.

37. Use a limited number of requests to attract the participants' attention to the presentations

The ECA should request the participant's attention maximum 2-3 times, with the number of requests lower for female participants. Then, it should rely on alternative strategies to attract attention (e.g., stop speaking for a few seconds).

38. Give participants time to become familiar with foreign names in a tour

Users should be given some time to become accustomed with names that are foreign to their cultural background. A short training session with the guide system prior the beginning of the tour will get users accustomed with names they may probably be hearing for the first time.

39. Prefer images of landmarks for navigation over other methods

Literature suggests that photographs of landmarks have a positive impact on the user's navigation ability. My experience suggests that an ECA improves this positive impact by augmenting the image with relevant and accurate verbal and non-verbal behaviours.

Simulation design

40. Use of high-definition sort-video clips over panoramas to simulate an outdoor environment

When testing in the field is not possible, use high-definition video clips to simulate an environment, over panoramas. Video-clips are easier and less expensive to produce. Furthermore, consider adding a treadmill to give users the "feeling" of walking from one location to another.

41. Consider the use of more "natural" methods for user interaction with a simulated environment

To avoid problems with users having difficulty synchronizing the mobile guide system with the simulated environment, consider the use of more natural methods of interaction (e.g., Microsoft Kinect)

- **Research method for user experience evaluation of ECA-based interfaces**

I designed and validated a method for an accessibility evaluation (discussed in experiment 6 of Chapter 8) of an ECA-based information presentation systems. This method combines data from face expression analysis, eye-tracking and retention tests to evaluate the accessibility of the content presented by an ECA presenter. As opposed to other advanced methods (e.g., measuring the user's brain activity), it provides high quality insights; it is cheaper and, most importantly, invisible to the user. Although studies measuring the attention-grabbing abilities of ECAs with eye-tracking have been reported in the past (e.g., Witkowsky *et al.* 2001), to my knowledge no method has been proposed that combines data from multiple sources to evaluate the accessibility of the content in ECA-based information presentation systems. This method is the most significant contribution of this research work to an open-area in ECA literature.

- **Personas**

The table below shows six sample personas, created based on my experimental studies and experiences, across three countries (Greece, UK, and USA). Although the personas are by not any means complete, they give interested readers an idea of the groups of people I encountered in my studies. I hope that other ECA researchers will find these personas useful.

Game user (GR)	First time user (GR)	Enthusiastic user (UK)	Young user/ UK	Empathetic user/USA	Older user/USA
					
PROFILE	***PROFILE***	***PROFILE***	***PROFILE***	***PROFILE***	***PROFILE***
1. Typical male, 30 years old **2.** Regular gamer (3-4 times / week). **3.** Likes the idea of a computer character used as a guide in an outdoor attraction. **4.** Would like an ECA with cutting-edge computer graphics and human-like behaviours. **5.** Does not like the use of text in mobile applications.	**1.** Typical female, 31 years old **2.** Focuses more on the content than how it is presented. **3.** An ECA may not impact her perception (positively or negatively) of the content. **4.** Prefers real human voice instead of T2S **5.** Needs well-structured content and grammatically correct	**1.** Typical male, 25 years old **2.** Typically spends more time exploring the content than any other user. **3.** Needs brief content, so he can continue exploring. **4.** Open to learn more about new cultures & religions. **5.** Excellent memory & attention skills.	**1.** Typical female, 25 years old **2.**Technologically savvy user in mobile technologies. **3.** Need to be visually involved in the content. **4.** Prefers an ECA with gestures and speech pauses. **5.** Open to learn more about new cultures and religions.	**1.** Typical female, 30 years old **2.** Highly empathetic user that can get involved quickly with an ECA. **3.** It assigns the ECA credibility to a point that she gets annoyed by its behaviour. **4.** She would most likely appreciate an ECA with humour. **5.** Needs to experience content where names unknown to her culture are properly explained.	**1.** Typical female, 60 years old **2.** Finds it difficult to follow ECA-based presentations. **3.** An ECA attracts too much of her attention. **4.** She would appreciate presentations of cultural value without an ECA. **5.** Finds a humorous attention-grabbing ECA distracting.

Scenario	*Scenario*	*Scenario*	*Scenario*	*Scenario*	*Scenario*
I like the idea of a computer character in the role of a guide in an attraction. A guide with "WOW" graphics would motivate me to pay more attention to the content and learn more about the castle. I do not like the use of text in mobile applications.	I like the idea of a mobile guide application, but the ECA does not affect me in any way. Regardless of the presentation method used, I like to experience content that is well-structured, grammatically correct, and is pronounced correctly by the system.	I like the idea of a Q&A system using natural language input. I need accurate answers to my questions fast and without having to rephrase my question many times. Furthermore, I would like content with simple vocabulary and sentence structure.	I prefer an ECA that makes me visually involved with the content, through the proper use of gestures and speech pauses, than an ECA without these attributes. In fact I believe that an ECA without these attributes makes it difficult to focus on the narrations of a mobile guide system.	I like an ECA that does not constantly requests my attention, and does not "yell" to me when I am not paying any attention. I like to experience content where the names of locations/artefacts unfamiliar to me or my culture are well explained.	I would like to experience less dense content over a long period of time. Furthermore, I do not like a humorous ECA, as its jokes distract me from the content.

FUTURE DIRECTIONS

The first avenue for future work is to repeat the experiments in the field, i.e., at the real castle of Monemvasia. I would like to improve the design of the mobile guide prototypes based on my current findings, update the systems to be compatible with current multi-touch hardware devices and run the experiments again. It may generate different results across all evaluation metrics. It will be interesting to compare the results generated by my current studies, with the results that will be produced by evaluating the prototype systems in the field.

Second, due to limited resources, the participants in my thesis studies were either university students or random people that have never visited the castle before. Utilizing real visitors of the castle may generate different results. For example, a real visitor that visits the castle for a full day (e.g., a history hobbyist) will most likely be more motivated to extract more information from the systems than the current participants. This factor could positively influence both subjective and objective evaluation metrics. Further, given that the castle of Monemvasia has hundreds of visitors every day, it should be fairly feasible to match gender and/or age across conditions in all of my experiments for cleaner experimental design.

Third, an interesting avenue for future studies is to adapt the prototypes to serve the needs of both long-term and short term-visitors of the castle. The visitors could be asked to complete a tour with the systems at their own convenience, without directly being observed by an experimenter. This would produce deeper insights on the user experiences, than those produced by the current studies.

Fourth, it will be interesting to validate my proposed method (see experiment six in Chapter 8) in the field. I would like to explore further the impact of the physical environment on my gathered data (e.g., fixations, heat maps, retention scores, etc.), what this data reveal about the ECA (e.g., whether or not it was effective in directing the users' attention focus on objects in his/her physical environment), and how these data compare with the data gathered from the study in the lab.

Fifth, another direction for future studies might be the development and evaluation of a fully interactive prototype of Talos – my authoring toolkit for ECA-based mobile guide applications. However, the task of implementing Talos is very challenging, most likely more suitable for a team of researchers than a single PhD student.

Finally, a mobile interface that could significantly impact the user's perception of an ECA is an augmented-reality interface. With such an interface, the ECA would be integrated into the physical environment, rather than merely using a background picture of the physical environment. Although the technical challenges of such an implementation are big, I believe is a suitable area where effort could be directed in future development and evaluation work of the systems.

References

Adams, R. (2005a): (R.G.Adams@mdx.ac.uk) (25 June 2005). RE: *"Accessibility Evaluation based on Simplex"* Personal email to I. Doumanis (idoumanis@mdx.ac.uk)

Adams, R. (2005b): *"Natural computing"*. In: Adams, R. (2005). Natural computing and interactive system design. London: Pearson, pp. 5-34.

Adams, R. (2005c): (R.G.Adams@mdx.ac.uk) (25 June 2005). RE: *"A questionnaire for user-interaction satisfaction"* Personal email to I. Doumanis (idoumanis@mdx.ac.uk)

Adams, R. and Langdon, P. (2003): *"SIMPLEX: a simple user check-model for Inclusive Design."* In: Universal Access in HCI: Inclusive Design in the Information Society. Stephanidis, C. (Ed.). 4, pp. 13-17. Mahwah: NJ: Lawrence Erlbaum Associates.

Adams, R. (2006): *"Decision and stress: cognition and e-accessibility in the information workplace.* Universal Access in the Information Society, 5 (4). pp. 363-379. ISSN 1615-5297

Adobe Inc., (2013). *"Adobe Flash"*. (Available from: http://www.adobe.com/products/flashplayer/) [Accessed: 19 February 2013]

Apple Inc., (2013). *"Siri Personal Assistant"*. (Available from: http://www.apple.com/ios/siri/) [Accessed February 19 2013].

Antonio Krüger, Jörg Baus, Dominik Heckmann, Michael Kruppa (2007): *"Adaptive Mobile Guides"*. In: The Adaptive Web, P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), Springer, 2007, pp. 521 - 549

Babu, S., Suma, E., Barnes, T., and Hodges, L.F. (2007): *"Can Immersive Virtual Humans teach Social Conversational Protocols?"* in: Proceedings of the IEEE International Conference on Virtual Reality 10 – 14 March 2007, Charlotte, N.C. pp. 215 - 218

Baylor, A. L. and S. Kim (2008): *"The Effects of Agent Nonverbal Communication on*

Procedural and Attitudinal Learning Outcomes". International Conference on Intelligent Virtual Agents. Tokyo, Springer**:** pp. 208-214.

Beun, R., de Vos, E. & Witteman, C. (2003): *"Embodied Conversational Agents: effects on memory performance and anthropormorphisation"*. In: Rist, T., Aylett, R., Ballin, D. & Rickel, J. (eds.) Intelligent Virtual Agents. Proceedings 4th International Workshop IVA2003 Kloster Irsee. LNAI 2792. Berlin: Springer, pp. 315-319.

Bickmore, T. & Picard, R. (2004): *"Establishing and Maintaining Long-Term Human-Computer Relationships"* Trans. on Computer-Human Interaction, Volume 12, Issue 2 (June 2005), pp. 293 - 327.

Bickmore, T. (2007): *"What Would Jiminy Cricket Do? Lessons From the First Social Wearable"*. In: Proceedings of the 2nd international conference on Online communities and social computing (2007), Berlin, Heidelberg, Springer, pp. 12-21

Bickmore, T. and Mauer, D. (2006): *"Modalities for Building Relationships with Handheld Computer Agents."* ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), April 22 – 27, 2006, Montréal, Canada, pp. 544 - 549

Bickmore, T., Schulman, D., Shaw, G. (2009): *"DTask & LiteBody: Open Source, Standards based Tools for Building Web-deployed Embodied Conversational Agents"*. Intelligent Virtual Agents, September 14 – 16, 2009, Amsterdam, Netherlands, pp. 425 - 431

Braun, N. (2003): *"Storytelling & conversation to improve the fun factor in software applications"*. In: Mark A. Blythe, Andrew F. Monk, Kees Overbeeke, and Peter C. Wright, editors, Funology, From Usability to Enjoyment, Dordrecht, April 2003. Kluwer Academic Publishers, pp. 233-241

Cassell, J., & Bickmore, T. (2000): *"External manifestations of Trustworthiness in the Interface"* Communications of the ACM, 43:12, pp. 50-56.

Cassell, J., & Stone, M. (1999): *"Living hand to mouth: Psychological theories about speech and gesture in interactive dialogue systems"*. Proceedings of the AAAI Fall symposium '99, pp. 34-42.

Cassell, J., Bickmore, T., Campbell, L. et al. (1999): *"Embodiment in Conversational Interfaces: Rea"* Proceedings of CHI '99, Pittsburgh, PA, pp. 520-527.

Catrambone, R., Stasko, J., & Xiao, J. (2002): *"Anthropomorphic agents as a user interface paradigm: Experimental findings and a framework for research"*, Proceedings of the 24th Annual Conference of the Cognitive Science Society, Fairfax, VA, August 2002, pp. 166-171

Chawla, P., Krauss, R. M., & Krieger, S. (1996): *"Conversational Visual Cues and Memory for Narrative"* (under editorial review).

Cheverst, K., Davies, N., Mitchell, K., Friday, A., & Efstratiou, C. (2000): *"Developing a Context-Aware Electronic Tourist Guide: Some Issues and Experiences"*. In: T. Turner, G. Szwillus, 206 References M. Czerwinski, & P. Fabio (Eds.), CHI 2000: Proceedings of the Conference on Human Factors in Computing Systems. The Hague, The Netherlands: ACM, pp. 17-24

Chin, D. (1991): "*Intelligent Interfaces as Agents*". In: Intelligent User Interfaces. J. Sullivan and S. Tyler (eds). ACM Press, New York, pp. 177-206.

Corbis, (1995): A Passion for Art – Renoir, Cézanne, Matisse and Dr. Barnes. CD-ROM. Corbis Publishing, 1995.

Cowell, A. J., & Stanney, K. M. (2002): *"User demographics for embodiment customization"*. AAAI Fall Symposium on Personalized Agent. Cambridge, MA: American Association for Artificial Intelligence. November 15-17, 2002, pp. n1-n2.

Cowell, A. J., & Stanney, K. M. (2003): *"On manipulating nonverbal interaction style to increase anthropomorphic computer character credibility"*. In Proc. of the AAMAS03 Workshop on Embodied Conversational Characters as Individuals, July 2003, Melbourne, Australia.

Cowell, A. J., Tanasse, T. E., & Stanney, K. M. (2003): *"Using anthropomorphic embodied conversational agents in mobile guides and information appliances"*. In proceedings of Workshop HCI in Mobile Guides, September 2003, Udine, Italy.

De Vos, Eveliene. (2002): *"Look at that Doggy in my Windows"*. Thesis (PhD).

Utrecht University.

Dehn, D. M., & Van Mulken, S. (2000): *"The impact of animated agents: a review of empirical research."* Int. J. Human-Computer Studies 52, pp. 1-22.

DELCA (2004): *"Enter the world of ghosts: New assisting and entertaining virtual agents*". Working paper, DELCA Ghost Project, IT University of Copenhagen. (Available from: http://www.itu.dk/research/delca/papers/delca_ghosts.pdf) [Accessed November 19 2005].

Dickerson, R., Johnsen, K., Raij, A., Lok, B., Hernandez, J., Stevens, A., and D. Lind. (2005): *"Evaluating a Script-Based Approach to Simulating Patient-Doctor Interaction"* SCS 2005, International Conference on Human-Computer Interface Advances for Modelling and Simulating (SIMCHI '05), 23-27 January 2005, New Orleans, pp.79-84.

eMarketView (2013) *"Eye tracking services"* (Available from: http://www.emarketview.com/servicios/usabilidad-conversion/eye-tracking) [Accessed February 19 2013].

Ellis, M. (2010, June 1). Streetmuseum: Q&A with Museum of London. In *Electronic Museum*. Consulted January 28, 2011. http://electronicmuseum.org.uk/2010/06/01/streetmuseum-qa-with-vicky-lee-museum-of-london/

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969): Pan-Cultural Elements In Facial Display Of Emotions Science, 164(3875), pp. 86-88.

Fraunhofer Institute for Integrated Circuits IIS, *"Sophisticated High-speed Object Recognition Engine (SHORE)"*, version 6.30, computer program and manual, Erlangen, Germany.

Franke, R. H. & Kaul, J. D. (1978). The Hawthorne experiments: First statistical interpretation. American Sociological Review, 1978, 43, pp. 623-643.

Finstad Kraif (2010): *"Response Interpolation and Scale Sensitivity: Evidence Against 5-point scales"* Journal of Usability Studies Vol. 5, Issue 3, May 2010, pp. 104-110.

Google Inc., (2013). *"Google Maps"*. (Available from: http://maps.google.co.uk/) [Accessed February 19 2013].

Guile3D studio, *"Virtual Assistant Denise"*, version 1.0, computer program and manual, Guile3D Studio, Brazil.

Glottopedia (2010) Thematic Role. [Online]. Available from: http://www.glottopedia.de/index.php/Thematic_role [Accessed 28/04/10].

Haraway, Donna. (1991): *"Simians, Cyborgs, and Women: The Reinvention of Nature"*. Routledge. New York.

Hile, H., Vedantham, R., Liu, A., Gelfand, N., Cuellar, G., Grzeszczuk, R., Borriello, G. (2008): *"Landmark-Based Pedestrian Navigation from Collections of Geotagged Photos"*. In: Proceedings of ACM International Conference on Mobile and Ubiquitous Multimedia (MUM 2008), December 3 – 5 2008, Umea, Sweden, pp.145-152.

Hollan, J., Hutchins, E., & Kirsh, D. (2000): *"Distributed Cognition: Toward a New Foundation for Human Computer Interaction Research"*. ACM Trans. on Computer-Human Interaction. 7 (2), pp. 174-196.

Hsu, J (1996): Multiple Comparisons: Theory and Methods, Chapman & Hall.

Institute for Creative Technologies (ICT), *"The ICT Virtual Human Toolkit"*, version 0.9.7.107, computer program and manual, University of Southern California, California, USA.

ISO 9241-11 (1998). Ergonomics requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability.

Johnson, W.L., LaBore, C., & Chiu, J. (2004): *"A pedagogical agent for psychosocial intervention on a handheld computer"*. AAAI Fall Symposium on Health Dialog Systems, October 22-24, 2004, pp. 22 – 24.

Joumana, M., *"Emotions and Face expressions"*. 2011. [online image] Available from: http://www.cedarseed.com/ [Accessed July 15 2011].

K Johnsen, D Beck, B Lok (2010): *"The Impact of a Mixed Reality Display Configuration on User Behaviour with a Virtual Human"*, 10th International Conference on Intelligent Virtual Agents (IVA 2010), September 20-22, 2010, Philadelphia, Pennsylvania, pp. 42-28.

Kaptelinin, Victor and Bonnie Nardi (1997): *"Activity Theory: Basic Concepts and Applications"* Proceedings of CHI '97, pp. 74–77.

Kjeldskov, J., Stage, J. (2004): *"New techniques for usability evaluation of mobile systems"* International Journal of Human-Computer Studies, Issue 60, pp. 599 – 620.

Krämer, N. C., Simons, N. & Kopp, S. (2007): *"The effects of an embodied agent´s nonverbal behavior on user´s evaluation and behavioural mimcry."* In C. Pelachaud et al. (eds.), *Intelligent Virtual Agents 2007 (pp. 238-251).*Berlin: Springer.

Kruppa, M., & Aslan, I. (2005): *"Parallel Presentations for Heterogenous User Groups – An Initial User Study"*. Proceedings of the first conference on intelligent technologies for interactive entertainment (INTETAIN 2005), November 30 – December 2, 2005, Madonna de Campiglio, Italy, pp. 54-63.

Kruppa, M., Lum, A., Niu, W., & Weinel, M. (2005): *"Towards mobile tour guides supporting collaborative learning in small groups"* PIA workshop in conjunction with User Modeling 2005, 24 – 25 July 2005, Edinburgh, UK, pp. 54 – 63.

Klein, D. and Manning, C. D. (2003): Accurate unlexicalized parsing. In ACL '03 Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, pp. 423 - 430.

Lang, A. (1995): *"Defining Audio/Video Redundancy from a Limited Capacity Information Processing Perspective"* Communication Research, 22, pp. 86-115.

Layar Inc., (2013) *"Layar Browser"* (Available from: http://www.layar.com/) [Accessed March 24 2013].

Lee, E.-J. & Nass, C. (1998): *"Does the ethnicity of a computer agent matter? An experimental comparison of human-computer interaction and computer-mediated communication"*. Proceedings of the Workshop on Embedded Conversational

Characters Conference. Lake Tahoe, CA, pp. 123-128.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., Bhogal, R. (1997): "*The persona effect: Affective Impact of Animated Pedagogical Agents*", Proc. CHI'97, pp. 359-366.

Likert, R. (1932): *"A Technique for the Measurement of Attitudes"*. In: Archives of Psychology 140, p.55

Lim, Y., Aylett, R. (2007): *"Feel the difference: a guide with attitude!"* In Pelachaud, C., Martin, J.C., Andre, E., Chollet, G., Karpouzis, K., Pele, D., eds.: Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA), Springer (2007) pp. 317 - 330

LookTel Inc., (2013) "*LookTel products*". (Available from: http://www.looktel.com/) [Accessed March 25 2013].

LearnLab (2013) Corrective self-explanation. [Online]. Available from: http://www.learnlab.org/research/wiki/index.php/Corrective_self-explanation. [Accessed 26/03/13]

Magnifis Inc., (2013). *"Robin Personal Assistant"*. (Available from: http://www.magnifis.com/wpress/) [Accessed March 26 2013].

MacDorman, K. F., & Ishiguro, H. (2006): *"The uncanny advantage of using androids in social and cognitive science research"*. Interaction Studies, Volume 7, Issue 3(2006), pp. 297–337.

MacDorman, K. F., Gadde, P., Ho, C.-C., Mitchell, W. J., Patel, H., Schermerhorn, P. W., & Scheutz, M. (2010): *"Probing people's attitudes and behaviours using humanlike agents."* Research Poster IUPUI Research Day. April 9, 2010. Indianapolis, Indiana.

Malaka, R. & Zipf, A. (2000): *"Deep map – Challenging it research in the framework of a tourist information system"*. In: Information and Communication Technologies in Tourism 2000. Proceedings of ENTER 2000, 7th. International Congress on Tourism and Communications Technologies in Tourism. Barcelona. Spain, 26 – 28 April 2000,

pp. 15-27

Massaro, D., Cohen, M., Beskow, J. *et al.* (2000): *"Developing and Evaluating Conversational Agents"* In: J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, (Eds.) Embodied Conversational Agents, pp. 287-318. MIT Press, Cambridge.

Mayer, R., & Moreno R. (2002): *"Animation as an Aid to Multimedia Learning."* Educational Psychology Review 14-1, pp. 87-99.

McBreen, H., & Jack, M., (2001): *"Evaluating Humanoid Synthetic Agents in E-Retail Applications",* IEEE Transactions on Systems, Man, and Cybernetics, Part A, Systems and Humans, vol.31, no.5, pp. 394-405

McBreen, H., Anderson, J., & Jack, M. (2001): *"Evaluating 3D Embodied Conversational Agents In Contrasting VRML Retail Applications"* Proc. Workshop on Multimodal Communication and Context in Embodied Agents, Autonomous Agent, June 2001, pp. 83-87.

Microsoft Inc., (2013) "Outlook 2013" (Available from: http://office.microsoft.com/en-gb/outlook/) [Accessed February 19 2013].

Morton, J., & Johnson, M. H. (1991): *"Conspec and Conlearn: a two-process theory of infant face recognition"*. Psychological Review, 98, pp. 164-181.

Miksatko, J., Kipp, K.H., Kipp, M. (2010): *"The Persona Zero-Effect: Evaluating virtual character benefits on a learning task."* In: Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA-2010), September 20-22, 2010, Philadelphia, PA, USA, pp. 475-482.

Moreno, R., Mayer, R. (2000): *"Engaging Students in Active Learning: The Case for Personalized Multimedia Messages."* Journal of Educational Psychology Vol 92(4), pp. 724-733.

Najjar, L. J. (1996): *"Multimedia Information and Learning"*. Journal of Educational Multimedia and Hypermedia, 5, pp. 129-150

Nass, C., Steuer, J., & Tauber, E. (1994): *"Computers are Social Actors"*. In:

Proceedings of the CHI Conference on Human Factors in Computing Systems: Celebrating Interdependence, April 24 - 28, 1994, Boston, MA, USA, pp. 72 – 78.

Nemetz, F., & Johnson, P. (1998): *"Towards Principled Multimedia"* [online]. University of London. Available from: http://www.bath.ac.uk/~mapfn/chi98/towards.htm [Accessed July 15 2011].

Ning Wang and Stacy Marsella (2006): *"Introducing EVG: An Emotion Evoking Game"* In: Proceedings of the 6th International conference on Intelligent Virtual Agents (IVA-2010), September 20-22, 2010, Philadelphia, PA, USA, pp. 282-291.

Norman, D. (1994): *"How might people interact with agents"* Communication of the ACM 37 (7), pp. 68-71.

Open Source Computer Vision (OpenCV) Wiki (2011) (Available from http://opencv.willowgarage.com/wiki/) [Accessed July 15 2011].

PandoraBots (2013). Available from: http://www.pandorabots.com/botmaster/en/home [Accessed: 19 February 2013]

Personality Forge Engine (2013). (Available from: http://www.personalityforge.com/) [Accessed: 19 February 2013]

Pospischil, G., Umlauft, M., & Michlmayr E. (2002): *"Designing LoL@, a Mobile Tourist Guide for UMTS"*. Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction, September 18-20, 2002, Pizza, Italy, pp.140-154

Proxem *"Advanced Natural Language Object oriented Processing Environment"*., version 0.8.7 computer program and manual, Proxem, France

QPC Inc., (2013) "Articulated Naturality Web" (Available from: http://www.qpcmobility.com/cn/index.html /) [Accessed March 24 2013].

Reeves, B., & Nass, C. (1996): *"The media equation: How people treat computers, television, and new media like real people and places"*. Cambridge University Press

Rickenberg, R., Reeves, B. (2000): *"The effects of Animated Characters on Anxiety,*

Task Performance, and Evaluations of User Interfaces". Proceedings of CHI 2000, The Hague, The Netherlands, Letters, 2(1), pp.49-56.

Rossen, B., Lind, S., Lok, B. (2009): *"Human-Centered Distributed Conversational Modeling: Efficient Modeling of Robust Virtual Human Conversations."* In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 474–481. Springer, Heidelberg (2009)

Schmorrow, D. D., & Kruse, A. A. (2004): *"Augmented Cognition"*. In: W. S. Bainbridge (Ed.), Berkshire Encyclopedia of Human-Computer Interaction. Great Barrington, MA: Berkshire Publishing Group, pp. 54-59.

Shneidermann, B., & Maes, P. (1997): *"Direct manipulations vs. interface agents: excerpts from debates at IUI'97 and CHI'97"*. Interactions, 4, pp. 42-61.

Simple Usability Inc., (2013) *"Emotion Response Analysis through EEG technique"* (Available from: http://www.simpleusability.com/) [Accessed February 19 2013].

Stocky, T., & Cassell, J. (2002): *"Shared Reality: Spatial Intelligence in Intuitive User Interfaces"*. In: IUI 2002: Proceedings of the Seventh International Conference on Intelligent User Interfaces, January 13 – 16, 2002, San Francisco, USA, pp. 224-225.

Stevens A, Hernandez J, Johnsen K, Dickerson R, Raij A, Jackson J, Min Shin, Cendan JC, Duerson M, Lok B, Lind DS (2006): *"The use of virtual patients to teach medical students communication skills"*. The American Journal of Surgery, Volume 191, Issue 6, pp. 806-811, June 2006

Sumby, W., & Pollack, I. (1954): *"Visual contribution to speech intelligibility in noise"* J. Acoustical Society America, vol. 26, no. 2, pp. 212-215.

Swartout, W., Traum, D., Artstein, R., Noren, D., et. Al (2010): *"Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides"*. 10th Int. Conf. on Intell. Virtual Agents, vol. 6353, Springer, Philadelphia, PA (2010), pp. 286-300

Smith, M., Lawrence, R. (2011): *"How to improve your memory."* [Online]. (Available from: http://www.helpguide.org/life/improving_memory.htm) [Accessed:

20th May 2011]

Takeuchi, A., & Naito, T. (1995): *"Situated Facial Displays: Towards Social Interaction."* Human Factors in Computing Systems: CHI'95 Conference Proceedings, May 7 - 11, 1995, Denver, Colorado, USA, pp. 450-455.

Tobias Eichner, Helmut Prendinger, Elisabeth André, and Mitsuru Ishizuka. *"Attentive presentation agents"* (2007): Proc 7th Int'l Conf on Int'l Virtual Agents (IVA'07), Springer LNCS 4722, Paris, France, Sept. 2007, pp. 283-295.

Travis, D. (2010): *"The 4 questions to ask in a cognitive walkthrough"*, weblog, (Available from: http://www.userfocus.co.uk/articles/cogwalk.html) [Accessed July 15 2010].

Van Mulken, S., André, E., & Muller, J. (1998): *"The Persona Effect: How substantial is it?"* In: Proceedings Human Computer Interaction (HCI-98), 1-4 September 1998, Sheffield, UK, pages 53-66.

Virtual People Factory (VPF) (2013) (Available from: http://www.virtualpeoplefactory.com) [Accessed March 11 2013].

Visual Recognition Inc., "eMotion" version 1.21 computer program and manual, Visual Recognition, Amsterdam, The Netherlands

Wagner, D. and D. Schmalstieg (2006): *Handheld Augmented Reality Displays*. In Proceedings of the 2006 IEEE conference on Virtual Reality (VR2006) Alexandria, Virginia, USA, March 25 – 29, 2006, pp. 321

Walker, J. H., Sproull, L., & Subramani, R. (1994): *"Using a Human Face in an Interface"*. In Proc. CHI 1994, April 24-28, 1994, Boston Massachusetts, USA, pp. 85-91.

Wallace, R. (2003). *"The Elements of AIML Style"*. [Online]. Available from: http://www.alicebot.org/style.pdf. [Accessed: 6/5/2010]

Wilson, M. (1997): *"Metaphor to personality: the role of animation in intelligent interface agents"*. Proceedings of the IJCAI-97, August 23-29, 1997, Nagoya, Aichi,

Japan, Workshop on Animated Interface Agents: Making them Intelligent.

Witkowski, M., Arafa, Y., & de Bruijn, O. (2001): *"Evaluating User Reaction to Character Agent Mediated Displays using Eye-tracking Equipment"*, In: Proc. AISB'01, Symp. on Information Agents for Electronic Commerce, 21 – 24 March 2001, University of York, UK, pp. 79-87

Wooldridge, M. (1999): "*Intelligent Agents*". In: Multi-agent Systems: A Modern Approach to Distributed Artificial Intelligence, ed. G. Weiss, pp. 27-77.Cambridge MA: MIT Press.

Wright, P., Milroy, R., & Lickorish, A. (1999): *"Static and animated graphics in learning from interactive texts"*. European Journal of Psychology of Education. Special issue on Visual Learning with New Technologies, XIV, pp. 203-224.

Xiao J., Stasko, J., & Catrambone, R. (2004): *"An Empirical Study of the Effects of Agent Competence on User Performance and Perception"*. Proceedings of AAMAS '04, New York, NY, July 2004, pp. 178-185

Xiao, J., Catrambone, R., & Stasko, J. (2003): *"Be Quiet? Evaluating Proactive and Reactive User Interface Assistants"*, Proceedings of INTERACT '03, Zurich, Switzerland, IOS Press, September 2003, pp. 383-390

Xiao, J., Stasko J., & Catrambone, R. (2002): *"Embodied Conversational Agents as a UI Paradigm: A Framework for Evaluation"* Proceedings of First International Joint Conference on Autonomous Agents & Multi-Agent Systems, AAMAS 2002, July 15 – 19 2002, Workshop on Embodied Conversational Agents: Lets specify and Compare Them!, Bologna, Italy.

Xiao, J. (2006): *"Empirical Studies on Embodied Conversational Agents"*. Thesis (PhD). Georgia Institute of Technology.

APPENDIX A.1:
Cognitive Walkthrough of the ICT Virtual Human and the Guile3D Toolkits

This appendix contains the cognitive walkthrough of the ICT Virtual Human and Guile3D toolkit. When appropriate and for practical purposes, I merge multiple trivial actions into solely meaningful user actions. In addition, only the actions that lead to potential usability problems are presented and discussed.

1) The ICT Virtual Human Toolkit Cognitive Walkthrough:

Task 1: Create a Question – Answering Dialogue

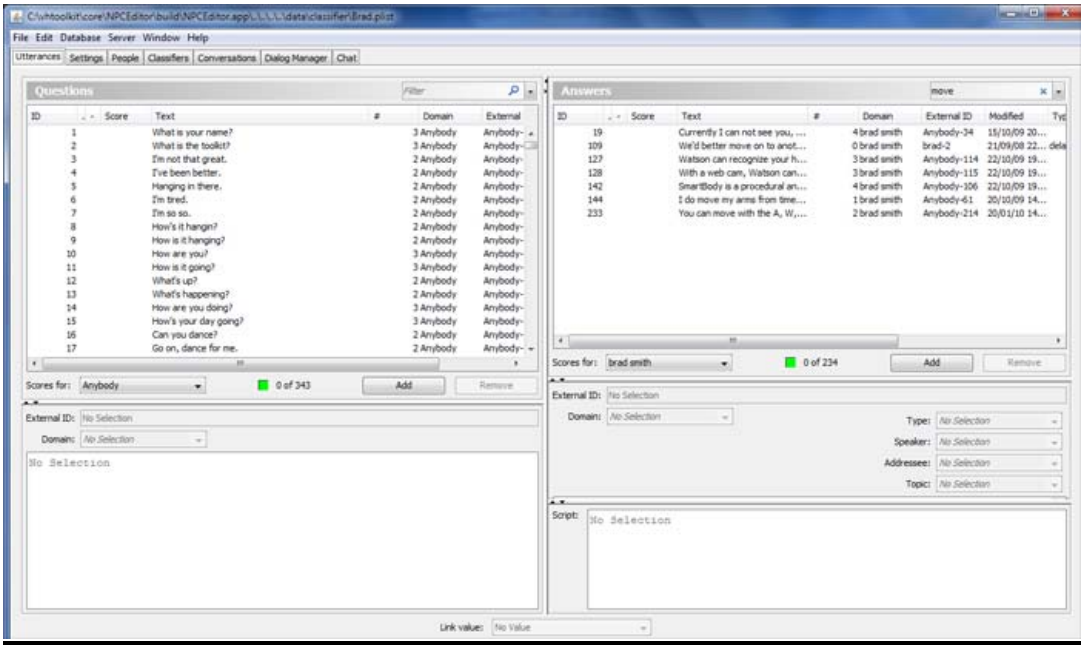


Figure A.1.1: The NPCEditor window (Source: ICT 2010)

For the purpose of the evaluation consider a simple dialogue with the following steps (taken from the prototypes):

- Initial phase of exchange of greetings.
- After this, the user states that s/he wants to start the tour. The agent begins describing a particular location.
- Once the agent's description is complete, the user states that s/he wants to ask questions about this location.

- The user asks a series of question and the dialogue returns to the initial phase.

The action sequence of this task is presented in terms of the user's action (UA) and the system's display or response (SD):

UA1: Create a new virtual agent, call it MGUIDE and save it.

SD1: The NPCEditor displays the full path of the saved agent on the user's hard disk drive.

UA2: Set the states of the example dialogue (as discussed above).

SD2: The display changes to the "People" panel.

UA3: Define the names of the dialogue states.

SD4: The state names are displayed in the "Name" panel on the left of the window.

UA5: Set the properties of the virtual agent.

- Move from state to state in the dialogue.
- Handle off topic utterances.
- Communicate with the rest of the modules in the toolkit.

SD5: Each property is shown correctly in the "Category" panel on the left of the window.

UA6: Map questions with answers for each state of the dialogue.

SD6: Display moves to "Utterances" panel.

UA7: Train the system to understand the question-answer pairs.

SD7: Display moves to "Classifiers" panel.

UA8: Test the question-answer pairs without having to start any other system module.

SD8: The display moves to the "Chat" window.

Task 1: Create a Question – Answering (QA) Dialogue

User action 2:

UA2: Set the states of the example dialogue (as discussed above)

SD2: The display changes to the "People" panel

Question 1: Will the users realistically try this action? Would the action occur to the user to do?

Yes, it is reasonable to assume that a researcher with basic dialogue modelling skills will do this as his/her first goal.

Question 2: Will the users see the control for the action? Is the control visible?

The control needed to set the dialogue steps is visible on the main interface of the editor

Question 3: Once users find the control, will they recognize that it is the one they want to complete the action?

No, it is not clear which tab is used to define the states of a dialogue. The tab named "Conversations" is a possible candidate, but it is used for another function. In fact the tab "People" is the correct choice, but it is quite possible that the user would fail at this point.

Question 4: Once the action has been taken is feedback appropriate, so users can go to the next action with confidence?

Yes, the display changes to the "People" panel.

User action 3:

UA3: Define the names of the dialogue states

SD4: The state names are displayed in the "Name" panel

Question 1: Will the users realistically try this action? Would the action occur to the user to do?

Yes, it is a safe assumption to make for the users of the toolkit.

Question 2: Will the users see the control for the action? Is the control visible?

Yes, the text fields needed to set the names of the states are visible in the window.

Question 3: Once users find the control, will they recognize that it is the one they want to complete the action?

No, it is difficult to associate the names of the text fields (i.e., "First Name:" and "Last Name :") to the user's goal at that point.

Question 4: Once the action has been taken is feedback appropriate, so users can go to the next action with confidence?

Yes, the name of the state is updated correctly on the "Name" panel.

User action 5:

UA5: Set the general properties of the virtual agent.

- Move from state to state in the dialogue.
- Handle utterances not related to the current topic of conversation.
- Communicate with the rest of the modules in the toolkit.

SD5: Each created property is shown correctly in the "Category" panel of the "Settings" tab.

Question 1: Will the users realistically try this action? Would the action occur to the user to do?

Yes, defining the general properties of the agent should be a goal for the user.

Question 2: Will the users see the control for the action? Is the control visible?

The "Settings" tab is visible in the main editor window.

Question 3: Once users find the control, will they recognize that it is the one they want to complete the action?

Yes, but s/he will not be able to fully complete the action. The control needed to connect the agent with the rest of the toolkit components is located on a different tab.

Question 4: Once the action has been taken is feedback appropriate, so users can go to the next action with confidence?

No, as there is no way to fully complete the action, feedback will be incomplete as well.

User action 7:

UA7: Train the system to understand the question-answer pairs

SD7: Display moves to "Classifiers" panel

Question 1: Will the users realistically try this action? Would the action occur to the user to do?

No, this action that should be performed automatically by the system during the process of editing the Question-Answer pairs in the "Utterances" panel.

Question 2: Will the users see the control for the action? Is the control visible?

Yes, the "Classifiers" tab is visible on the main window of the editor.

Question 3: Once users find the control, will they recognize that it is the one they want to complete the action?

No, the Classifiers window has several jargon terms that are impossible to understand even for very advanced users.

Question 4: Once the action has been taken is feedback appropriate, so users can go to the next action with confidence?

No, feedback is not returned in plain English, but rather in system parameters.

Task 2: Add character gesticulation/facial expressions to the responses

The toolkit selects and synchronises automatically the character's gesticulation. It relies on hand-crafted rules, but it doesn't provide any interface to aid their creation. Therefore, it is not possible to conduct a cognitive walkthrough for this task.

Task 3: Add multimodal input

Assume that in the initial phase of the dialogue, I want the character to say "Hello" first, by reacting to the presence of the user:

UA1: Start Watson from the Launcher window.
SD1: The Watson Stereo Tracker window appears.
UA2: Start the NPCEditor from the Launcher window.
SD2: The NPCEditor Appears.
UA3: Create an agent property named "User Recognition".
SD3: Display moves to the "Settings" window of the NPCEditor.
UA4: Create a label called Presence
SD4: The label is shown correctly in the Token panel at the bottom of the window
UA5: Add code to map computer vision messages to the "presence" property
SD5: Display moves to "Dialog Manager" window

User action 5:

UA5: Add code to map vision messages to the "presence" label
SD5: Display moves to "Dialog Manager" window.

Question 1: Will the users realistically try this action? Would the action occur to the user to do?

No, users would expect that the necessary labels (and their backend code) are

available by default to the system.

Question 2: Will the users see the control for the action? Is the control visible?

Yes, the "Dialogue Manager" tab is visible on the interface of the NPCEditor

Question 3: Once users find the control, will they recognize that it is the one they want to complete the action?

Yes, it is a safe assumption to make for the users of the toolkit.

Question 4: Once the action has been taken is feedback appropriate, so users can go to the next action with confidence?

No, the code editor doesn't provide any feedback on the range of functions supported by the toolkit's integrated computer vision module (aka Watson).

2) The Guile3D Cognitive Walkthrough

Task 1: Create a Question – Answering Dialogue

The same dialogue will be considered in this section:

- Initial phase of exchange of greetings.
- After this, the user states that s/he wants to start the tour. The agent begins describing a particular location.
- Once the agent's description is complete, the user states that s/he wants to ask questions about this location.
- The user asks a series of question and the dialogue returns to the initial phase.

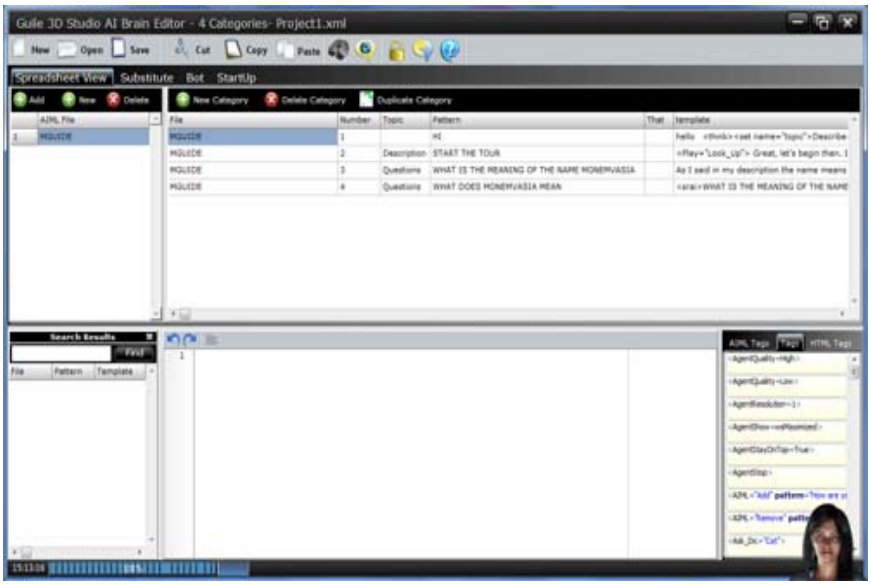


Figure A.1.2: The AI editor (Source: Guile3D 2010)

UA1: Create a new AIML file, call it "MGUIDE" and save it.

SD1: The filename appears in a panel on the right of the window.

UA2: Set the states of the example dialogue (as discussed above).

SD2: The names of the dialogue states appear in the topic text-field.

UA3: Map questions with answers for each state of the dialogue.

SD3: Patterns and templates appear correctly in the relevant fields.

UA4: Save the AIML file.

SD4: A popup window appears that indicate the success of the operation.

UA5: Add the necessary animation tags for controlling Denise's face expressions.

SD5: The animation tags are shown within the template's text.

UA6: Compile the AIML file.

SD6: A message window appears to save the Encrypted AIML files.

UA7: Load the files by restarting the application.

SD7: Denise loads and replies to our questions correctly.

Task 1: To create a Question – Answering (Q&A) Dialogue

User Action 5:

UA5: Compile the AIML file.

SD5: A message window appears to save the Encrypted AIML files.

Question 1: Will the users realistically try this action? Would the action occur to the user to do?

No, users would expect that "Save", would save the AIML file in the proper format for the dialogue engine.

Question 2: Will the users see the control for the action? Is the control visible?

Yes, the "Compile" button is visible on the Editor's menu bar.

Question 3: Once users find the control, will they recognize that it is the one they want to complete the action?

Yes, it is a safe assumption to make for the users of the toolkit.

Question 4: Once the action has been taken is feedback appropriate, so users can go to the next action with confidence?

Yes, the toolkit provides feedback that the operation was completed in both text and audio form.

Task 2: Add character gesticulation/facial expressions to the responses

User Action 5:

UA5: Add animation tags for controlling Denise's face expressions.

SD5: The animation tags are shown within the text assigned as an answer.

Question 1: Will the users realistically try this action? Would the action occur to the user to do?

Yes, it is reasonable to assume that a researcher with basic dialogue modelling skills will do this as his/her first goal.

Question 2: Will the users see the control for the action? Is the control visible?

Although there is a tag tab with several tags, it is not clear which of these tags are designed for animating Denise's face. I entered the animation tags manually.

Question 3: Once users find the control, will they recognize that it is the one they want to complete the action?

No, the tags are laid out alphabetically and not according to category.

Question 4: Once the action has been taken is feedback appropriate, so users can go to the next action with confidence?

No there is no feedback (visual or otherwise) to indicate what these tags actually do.

Task 3: Add multimodal input

Although Denise has an advanced face recognition module, that performs various functions; it has not yet been made available with the toolkit. Therefore, it was not evaluated.

Appendix A.2: Talos prototypes

Creating and handling the necessary data for the prototypes is a tedious and complex process. Therefore, inspired by the architecture of Talos two simple tools were developed: a) a simple Artificial Intelligence editor to aid the various development tasks (e.g., to build and edit databases with questions and answers, character scripts and props etc.) and b) a script parser.

1) A simple Artificial Intelligence editor

In particular this tool enables the designer to:

1) Develop questions-answers knowledge bases in XML and AIML format.
2) Automatically translate question-answer sets into other languages (e.g., Greek, French, etc.).
3) Edit various scene and avatar attributes (e.g., various props like hats, hair, etc., scene backgrounds and others).
4) Build databases with location-sensitive scripts representing a route. Tests the scripts in real-time for location accuracy.
5) Build Automatic Speech Recognition (ASR) grammars for a variety of speech recognition engines and test them.

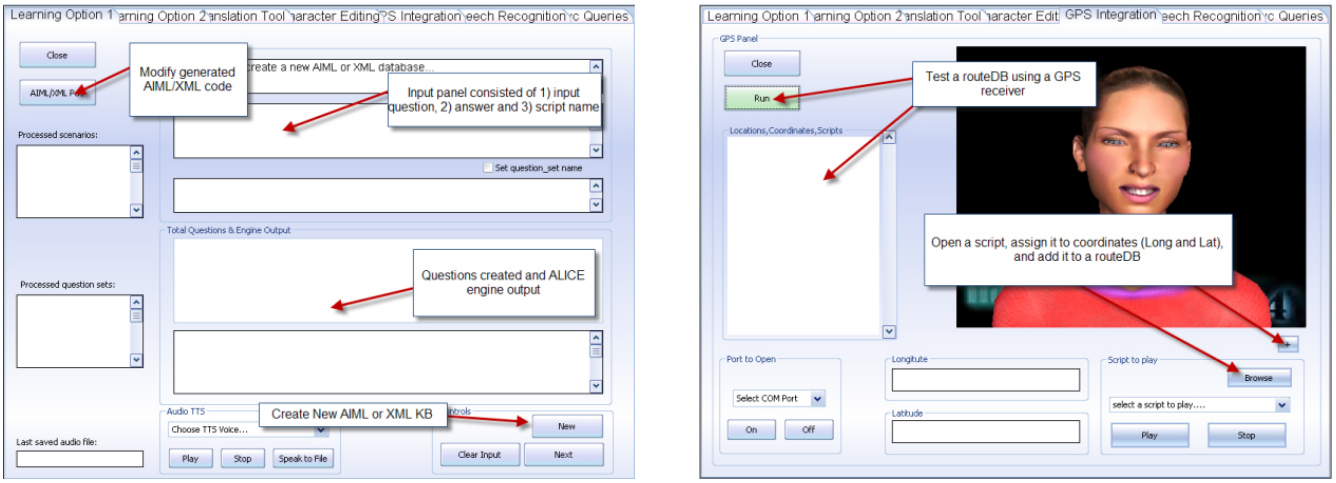


Figure A.2.1: Screenshots of the UI editor

2) A Haptek Script Parser

The avatar engine used in the current implementation of the prototypes is from Haptek[1] Corp. Developers have absolute control over the engine's output using text commands stored in scripts. However, to execute character actions in the right order (e.g., to synchronize character gesticulation or other scene action, with the spoken text) these commands must be timed precisely. As Haptek doesn't provide any tool to automate this process, each script must be created and timed manually.

[1] The Haptek corporate homepage at: http://www.haptek.com

\book =<back,C17> Please \book=anim,back> head your way back to the main street of the castle. After making \book=anim,portello6> at the first opportunity two left turns……

Figure A.2.2: Tagged text as input for the script parser

To address this issue, I created a script parser that takes tagged text input and automatically creates the proper scripts that correctly synchronise scene action with the spoken audio. For example from the tagged text in Figure A.2.2, the parser creates a script that correctly synchronizes the character gestures (named “back” and “portello6”) with the spoken audio and changes the background of the scene once the script is loaded.

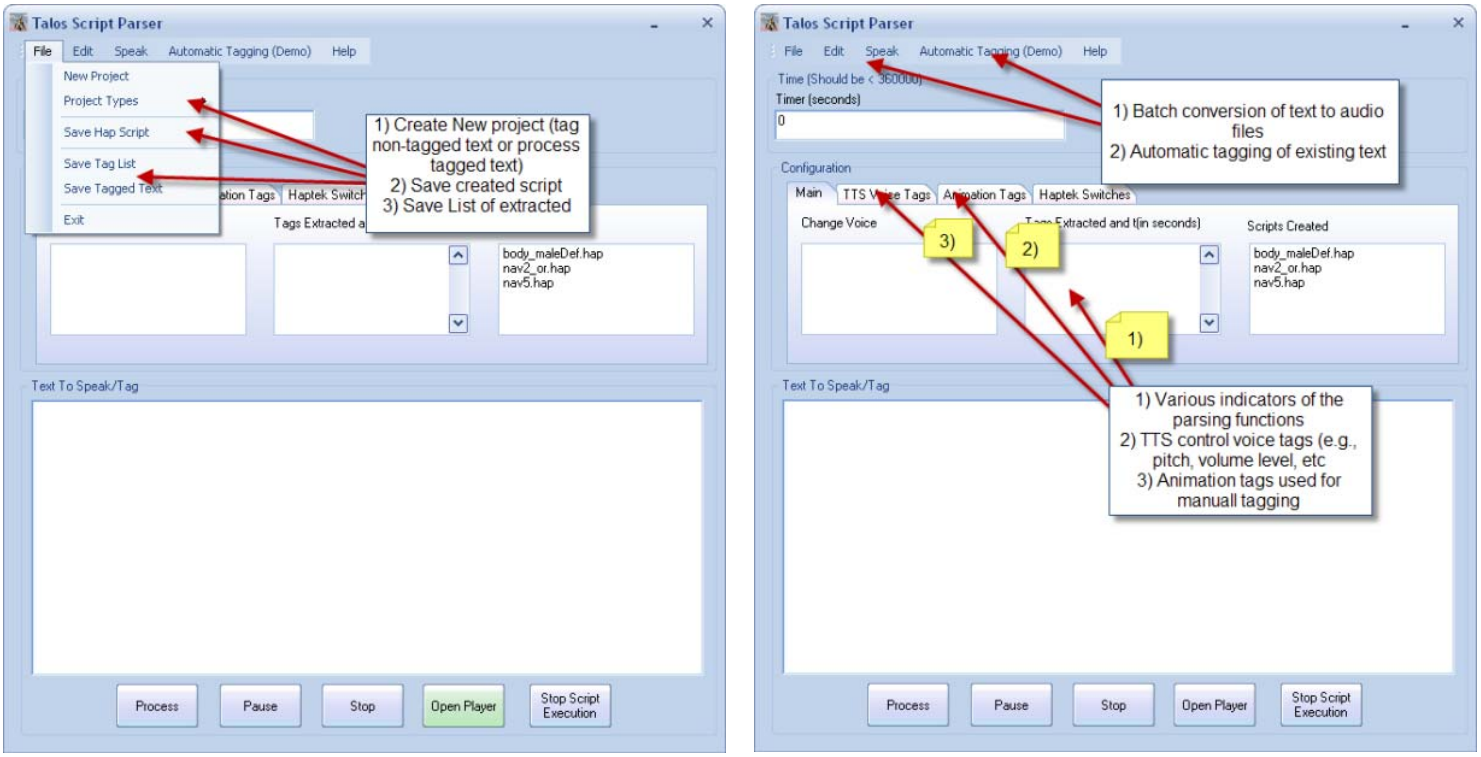


Figure A.2.3: Screenshots of the script parser

To further aid the creation of Haptek scripts, the parser provides the following functions:

- Reuse control tags extracted from previous scripts.

The tool stores every tag (shown in Figure A.2.2 in red) extracted from texts in libraries. These tags can be used to tag texts in any language.

- Limited support for automatically tagging texts

The tool has a limited support for automatic tagging of arbitrary texts. In the current implementation, it works (within limits) in the domain of MGUIDE, but it can be adapted to match the requirements of any domain with minimal effort.

- Text to Speech (TTS) Control tags

Developers can add T2S action tags in the text to control various features of the text-to-speech synthesis (e.g., set up the volume, rate and pitch for all text to be read, insert human sounds like laughter, take a breath, cough, and others etc.)

- Single/batch conversion of any text to the audio format (i.e., .ogg) used by the Haptek engine.

A sample script generated by the parser:

```
#Haptek  Version= 2.00 Name=Nav2E_A.hap HapType= script FileType= text
## world It
##prereq= none

\clock [t= 0.0] \load [file= [Nav2E_A.ogg]]
\clock [t=0] \SetSwitch [figure= fullBod switch= wide_stair1 state= start]
\clock [t= 2]\loadbackgrnd [file= [landmrk6.jpg]]
\clock [t= 5] \SetSwitch [figure= fullBod switch= wide_stair2 state= start]
```

Figure A.2.4: A sample script generated by the parser

Appendix B:

The information contained in this appendix relates to the development of the match and search algorithm. The algorithm is discussed in detail in chapter 6 (see §6.5).

```
        '1. Tag the user's input
            Dim input As IList(Of IWord) = tagger.TagText(userinput3)
        For Each token_input As IWord In input
            If common_POS.Contains(token_input.PartOfSpeech.ToString) Then
            Else
                input_list.Add(New KeyValuePair(Of String,
String)(token_input.Text, token_input.PartOfSpeech.ToString))
            End If
        Next
        '2. Tag the VPF trigger
            Dim trigger As IList(Of IWord) = tagger.TagText(VPF_trigger)
        For Each VPF As IWord In trigger
            If common_POS.Contains(VPF.PartOfSpeech.ToString) Then
            Else
                trigger_list.Add(New KeyValuePair(Of String, String)(VPF.Text,
VPF.PartOfSpeech.ToString))
            End If
        Next
        '3. Filter input and Trigger based on the list of keywords provided by
the VPF service.
        Dim xmldoc As New XmlDocument

xmldoc.Load("http://vpf.cise.ufl.edu/VirtualPeopleFactory/virtual_patient_mvc/V
iew/web_service.php?model=Script&primary_key_value=" & Current_Script &
"&method=getGlobalKeywords&encoding=xml&username=giannis&password=wgb145")
        Dim nodeList As XmlNodeList =
xmldoc.DocumentElement.SelectNodes("array_item")
        Dim nodes As New List(Of String)

        For Each node As XmlNode In nodeList

global_keywords.Add(node.SelectSingleNode("global_keyword_text").InnerText)
        Next
        'Filter the input first
        For Each keyword As String In global_keywords
            For index1 As Integer = 0 To input_list.Count - 1
                If input_list.Item(index1).Key = keyword Then
                    new_input_list.Add(New KeyValuePair(Of String,
String)(keyword, input_list.Item(index1).Value))
                Else
                End If
            Next
        Next
        'Filter the trigger second
        For Each keyword As String In global_keywords
            For index2 As Integer = 0 To trigger_list.Count - 1
                If trigger_list.Item(index2).Key = keyword Then
                    new_trigger_list.Add(New KeyValuePair(Of String,
String)(keyword, trigger_list.Item(index2).Value))
                Else
                End If
            Next
        Next
```

```
        TextBox1.Text = ""
        For index As Integer = 0 To new_input_list.Count - 1
            TextBox1.Text = TextBox1.Text & Space(1) &
new_input_list.Item(index).Key.ToString
        Next
        '4. Compare what is left for POS and values.
        '4.1 If input.count < Triggers.count then compare tokens with or
without the same value for different POS
        If new_input_list.Count <> 0 Then
            If new_input_list.Count < new_trigger_list.Count Then
                For index1 As Integer = 0 To new_trigger_list.Count - 1
                    If index1 < new_input_list.Count Then
                        If (new_input_list.Item(index1).Key <>
new_trigger_list.Item(index1).Key Or _
                            new_input_list.Item(index1).Key =
new_trigger_list.Item(index1).Key) And _
                        (new_input_list.Item(index1).Value <>
new_trigger_list.Item(index1).Value) Then
                            Failed_Comparisons.Add("Failed")
                        Else
                            If (new_input_list.Item(index1).Key =
new_trigger_list.Item(index1).Key And _
                                new_input_list.Item(index1).Value =
new_trigger_list.Item(index1).Value) Then
                                Success_Comparisons.Add("Succesfull")
                            End If
                        End If
                    Else
                        If (new_input_list.Item(new_input_list.Count - 1).Key
<> new_trigger_list.Item(index1).Key Or _
                            new_input_list.Item(new_input_list.Count - 1).Key =
new_trigger_list.Item(index1).Key) And _
                        (new_input_list.Item(new_input_list.Count - 1).Value <>
new_trigger_list.Item(index1).Value) Then
                            Failed_Comparisons.Add("Failed")
                        Else
                            If (new_input_list.Item(new_input_list.Count -
1).Key = new_trigger_list.Item(index1).Key And _
                                new_input_list.Item(new_input_list.Count -
1).Value = new_trigger_list.Item(index1).Value) Then
                                Success_Comparisons.Add("Succesfull")
                            End If
                        End If
                    End If
                Next
            End If
            '4.2 If input.count = triggers.count then compare tokens for values
only
            If new_input_list.Count = new_trigger_list.Count Then
                For index1 As Integer = 0 To new_trigger_list.Count - 1
                    If (new_input_list.Item(index1).Key <>
new_trigger_list.Item(index1).Key Or _
                        new_input_list.Item(index1).Key =
new_trigger_list.Item(index1).Key) And _
                    (new_input_list.Item(index1).Value <>
new_trigger_list.Item(index1).Value) Then
                        Failed_Comparisons.Add("Failed")
                    Else
                        If (new_input_list.Item(index1).Key =
new_trigger_list.Item(index1).Key And _
                            new_input_list.Item(index1).Value =
new_trigger_list.Item(index1).Value) Then
```

```
                            Success_Comparisons.Add("Succesfull")
                        End If
                    End If
                Next
            End If
            '4.3 If input.count > triggers.count then
            If new_input_list.Count > new_trigger_list.Count Then
                For index1 As Integer = 0 To new_input_list.Count - 1
                    If index1 < new_trigger_list.Count Then
                        If (new_input_list.Item(index1).Key <>
new_trigger_list.Item(index1).Key Or _
                            new_input_list.Item(index1).Key =
new_trigger_list.Item(index1).Key) And _
                        (new_input_list.Item(index1).Value <>
new_trigger_list.Item(index1).Value) Then
                            Failed_Comparisons.Add("Failed")
                        Else
                            If (new_input_list.Item(index1).Key =
new_trigger_list.Item(index1).Key And _
                                new_input_list.Item(index1).Value =
new_trigger_list.Item(index1).Value) Then
                                Success_Comparisons.Add("Succesfull")
                            End If
                        End If
                    Else
                        If (new_input_list.Item(index1).Key <>
new_trigger_list.Item(new_trigger_list.Count - 1).Key Or _
                            new_input_list.Item(index1).Key =
new_trigger_list.Item(new_trigger_list.Count - 1).Key) And _
                        (new_input_list.Item(index1).Value <>
new_trigger_list.Item(new_trigger_list.Count - 1).Value) Then
                            Failed_Comparisons.Add("Failed")
                        Else
                            If (new_input_list.Item(index1).Key =
new_trigger_list.Item(new_trigger_list.Count - 1).Key And _
                                new_input_list.Item(index1).Value =
new_trigger_list.Item(new_trigger_list.Count - 1).Value) Then
                                Success_Comparisons.Add("Succesfull")
                            End If
                        End If
                    End If
                Next
            End If
            If Success_Comparisons.Count = Failed_Comparisons.Count Then
                comparison = "Succesfull"
            ElseIf Success_Comparisons.Count > Failed_Comparisons.Count Then
                comparison = "Succesfull"
            ElseIf Success_Comparisons.Count < Failed_Comparisons.Count Then
                comparison = "Failed"
            End If
        End If
        If (comparison = "Failed" Or comparison = "") Then
            'pass the input for predicate analysis
            predicate_test(userinput3)
        End If
```

Figure B.1: Snippet of the search and match algorithm

```
<?xml version="1.0" encoding="utf-8"?>

<Location_A>
<sentance id="1">
<text>Let us begin the tour then</text>
<predicates>begin</predicates>
<Deep_Syntax name="Subject">Us</Deep_Syntax>
<Deep_Syntax name="DirectObject">Tour</Deep_Syntax>
</sentance>

<sentance id="2">
<text>I am ready let us begin</text>
<predicates>begin</predicates>
<Deep_Syntax name="Subject">Us</Deep_Syntax>
 </sentance>

<sentances id="3">
<text>Can we begin the tour please</text>
<predicates2>begin</predicates2>
<predicates2>please</predicates2>
<Deep_Syntax name="Subject">We</Deep_Syntax>
<Deep_Syntax name="Subject">Tour</Deep_Syntax>
</sentances>

<sentance id="4">
<text>Let us go then</text>
<predicates>go</predicates>
<Deep_Syntax name="Subject">Us</Deep_Syntax>
</sentance>

<sentances id="5">
<text>Does the castle has any other gates</text>
<predicates2>do</predicates2>
<predicates2>have</predicates2>
<Deep_Syntax name="Subject">Castle</Deep_Syntax>
<Deep_Syntax name="DirectObject">Gates</Deep_Syntax>
</sentances>

<sentance id="6">
<text>I want more information about the main gate of the Upper Town</text>
<predicates>want</predicates>
<Deep_Syntax name="Subject">I</Deep_Syntax>
<Deep_Syntax name="DirectObject">Information</Deep_Syntax>
<Deep_Syntax name="PrepObject">Main-Gate</Deep_Syntax>
</sentance>

<sentance id="7">
<text>I want to listen about gate 1 the main gate of the Upper Town</text>
<predicates>listen</predicates>
<Deep_Syntax name="Subject">I</Deep_Syntax>
<Deep_Syntax name="PrepObject">Main-Gate</Deep_Syntax>
</sentance>

<sentance id="8">
<text>I want to listen about gate 2 the Portello</text>
<predicates>listen</predicates>
<Deep_Syntax name="Subject">I</Deep_Syntax>
<Deep_Syntax name="DirectObject">Portello</Deep_Syntax>
</sentance>
```

```
<sentance id="9">
<text>I want more information about the Portello</text>
<predicates>want</predicates>
<Deep_Syntax name="Subject">I</Deep_Syntax>
<Deep_Syntax name="DirectObject">Information</Deep_Syntax>
<Deep_Syntax name="PrepObject">Portello</Deep_Syntax>
</sentance>

<sentance id="10">
<text>I want to listen more about gate 3 the west gate of the citadel</text>
<predicates>listen</predicates>
<Deep_Syntax name="Subject">I</Deep_Syntax>
<Deep_Syntax name="DirectObject">West-Gate</Deep_Syntax>
</sentance>

<sentance id="11">
<text>I want more information about the west gate of the citadel</text>
<predicates>want</predicates>
<Deep_Syntax name="Subject">I</Deep_Syntax>
<Deep_Syntax name="DirectObject">Information</Deep_Syntax>
<Deep_Syntax name="PrepObject">West-Gate</Deep_Syntax>
</sentance>

</Location_A>
```

Figure B.2: Excerpt from the XML database the algorithm uses

APPENDIX C:

The information contained in this appendix relates to an exploratory study I conducted in the actual castle of Monemvasia. The study was presented at a conference in Austria.

Humanoid Animated Agents in Mobile Applications: An Initial User Study and a Framework for Research

Ioannis Doumanis, Ray Adams, Serengul Smith

Abstract Research on humanoid animated agents for mobile guide systems, has paid insufficient attention to the evaluation of such interfaces. In addition, the few existing studies, suffer from the absence of a rigid framework, in which detailed research can be conducted and detailed findings subsumed. In this paper we propose a theoretically well-supported framework, consisted of four key components that can systematize the research: differences in users, the agents, the mobile environments, and the task the user is performing. Our first experiment, within this framework, manipulated the agent's presence (present versus absent) and order of presentation (present versus absent and vice versa). We found that, the user's experience was influenced by the agent's visual presence, with this effect interacting with the order of presentation; while practice resulted from the order manipulation affected the perception of the visual agent and the objective time performance. Finally, where appropriate, we used these findings to derive a number of hypotheses, on what to expect from future experiments.

Keywords Animated agents, mobile interfaces, evaluation, empirical study

1 Introduction

Over the last 10 years, the world has seen a tremendous progress in mobile technologies, with high-bandwidth wireless networks becoming more pervasive and mobile devices becoming progressively smaller and smarter. While problems of content delivery and storage are well on the way to be solved, the issue of how the user should interact with these devices is still being debated. Currently, the user interface of these devices is based on a variation of the "traditional" graphical user interface (GUI) for desktop computers. However, unlike typical stationary computer scenarios, where the context is more or less static, in mobile scenarios the dynamic nature of the user's situation rapidly affects his ability to process, store and respond to information. Given this fact, and the increased complexity of mobile applications, and thereby, of their underlying GUI's that support them, the direct manipulation metaphor becomes a bottleneck, in the accessibility and usability of these systems.

The use of animated agents has been recently suggested [6] as a more natural and transparent interface that can potentially solve this problem. Conversational interfaces appear as a more natural interactional style, because the user doesn't have to learn complex command structure and functionality, to operate a system [11]. In addition, an animated agent could use body (e.g., intonation, and gestures) to not only supplement, but also to complement the information conveyed in language. Such agents have already been developed for mobile applications (e.g., [31), and several more projects are well underway for more technological advancements. However, insufficient attention has been paid to the empirical evaluation of such interfaces. The direct consequence is that there is truly a near absence of evidence, on the potential effects of animated agents on the users of mobile applications. In the animated agent research for stationary systems, some effects on the user's attitudes and cognition, have already been established, but, relating those effects to the human user in mobile environments, is yet to be done. Furthermore, the few existing studies lack of a common and theoretically well-supported frame -work, to guide their evaluations. As a result of this scarce knowledge, there are risks in utilizing animated characters in mobile interfaces. If the animated agent does not actually enhance the system, or it is not appropriate for the particular scenario, the user may become distracted and the entire interaction may collapse. Then, there is the concern, that this research could build up results, that are inconsistent or equivocal - currently a common phenomenon, among the empirical evaluations of stationary applications [8]): - thus, limiting the already limiting consensus, on the precise effects of the various aspects of animated agents, on the mobile human user.

Our goal is to develop a comprehensive framework that is theoretically well supported, in order to systematically evaluate and understand the animated agent as a user interface paradigm, in the context of electronic mobile tour guide applications. The present paper outlines the framework and an experiment that examines the most fundamental question within this framework, whether the mere presence of an animated agent on a mobile interface, has some kind of measurable effect to the human user.

2 The underlying theory

Investigating the potential effects of animated agents to the user in mobile contexts, requires first, and foremost, the establishment of a sound theoretical foundation, for supporting empirical evaluations of users and mobile systems/animated agents, and interpreting the results. We propose the use of Simplex Two model of human cognition, and the embodied tenet of distributed cognition as a theoretical basis, for empirical study in this area.

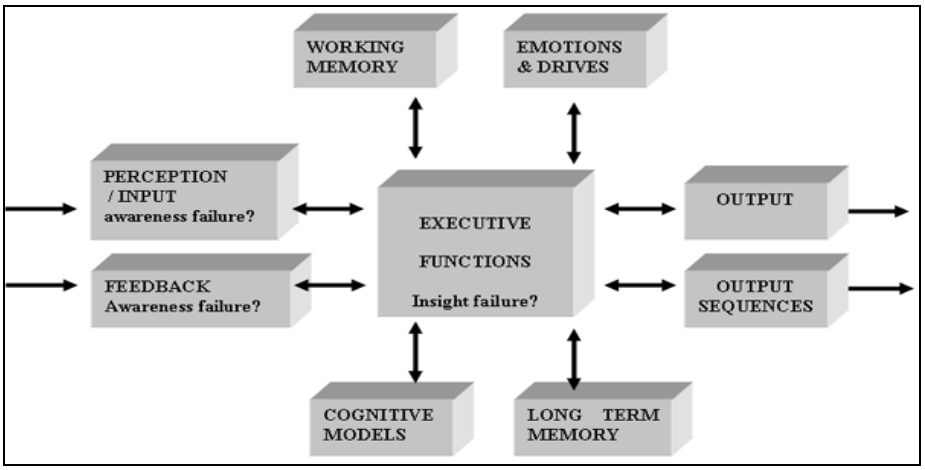


Fig.1 A Depiction of Simplex Two (source [1])

The Simplex Two model [1] postulates nine (see Fig. 1) zones or modules (validated by [2]) of intelligent human behaviour, each of which can act partially independently of each other. The current use of Simplex is that of underpinning measures of performance and accessibility of mobile guide systems, featuring anthropomorphic animated agents, and supporting a systematic evaluation of the user's psychology. Although, the Simplex Two model provides a generative view of the human cognition, it tends to focus more on the cognition of the individual. The nature of a mobile situation, however, requires more ubiquitous cognitive processes from the user. Thus, there is a need to postulate an additional component to Simplex Two, in order to encompass interactions, between people with objects and structures, in the physical environment. This component is the embodied tenet, of distributed cognition, and it is discussed below:

2.1 Embodied tenet of distributed cognition

The central tenet [9] of the distributed cognition approach is that cognition is embodied. It is not incidental that we have bodies and we use them for causally linking with our immediate environments. This relation is an essential fact of cognition that evolution has designed us to exploit. In other words, this approach postulates that the human mind is not a passive representational engine whose only function is to create internal models of the external world. There is a closer and far more complex relation between internal and external processes, which involves coordination at many different time scales, between internal resources such as memory and attention, and external resources such as the objects and artefacts, constantly surrounding us. A crucial moment within this coordination, that can decide its success, is whenever the user needs to switch his attentional focus from the virtual world of the mobile device to the actual objects/artefacts of the physical world. An animated agent could be used, to effectively guide the user's attentional focus towards these physical objects, by helping the user in understanding the underlying structure in the physical space. It may for example, help the user to locate a particular landmark by directing his gaze towards its current location in the virtual world, and hence allow him to better locate it in the physical environment. In the very same way, the

animated agent may help the user to identify parts of the current structure/object he is looking at, and for which the system is providing information for.

3 A framework for research on anthropomorphic animated agents for mobile interfaces

Once the theoretical foundation has been presented, the next step is to consider the key factors that can influence the accessibility and usability of such interfaces. For this reason, we suggest an investigative framework for studying animated agents in mobile interfaces, consisted of four key components: characteristics of the user; the agent; the mobile environment in which the user is performing tasks; and the task the user is performing. This framework draws on a number of previous works [7], [14], and theories discussed above. Below, we first discuss a number of variables that are possible within each factor and then conclude with those, that we are interested to test, as part of the present research programme.

Factor 1: Features of the User

Potential users, vary of course, in many ways. However, based on options of Simplex Two and the embodied tenet of distributed cognition, we can derive certain features that may be quite likely to affect, how accessible and/or usable a user finds a mobile interface with an animated agent. These features include:

Gender: The user's gender can play a role in the overall perception of the agent's characteristics. For instance, [12] showed, that for an electronic commerce desktop application, female participants deemed a set of female cartoon-like agents to be more polite, than the corresponding males, and male participants thought the male cartoon-like agents to be more polite than the female ones.

Perceptual capability: The ability of a user, to perceive the information provided by an animated agent in a mobile setting, may vary greatly between individuals. Although, most people use a combination of all three perceptual models, sight sound and touch, there are certain types of individuals, who strongly prefer a particular modality and have trouble in using other communication channels. For those users, the multi-modal nature of the human-agent communication can be proven to be problematic. For example, an auditory user that communicates best by hearing and verbalizing would most likely appreciate au animated agent with excellent verbal output, and minimal non-verbal presence. Conversely, a visual user who processes information mostly by seeing and verbalizing, would most likely give the non-verbal channels a higher priority in his communications with the animated agent than the verbal channels.

Age: The above factors can be differentiated greatly, among users of different age groups. The users of older age groups generally have declined perceptual and cognitive abilities, as well as motor responses, compared to those in the younger age group. Therefore, it is reasonable to expect, that

the attitudes and performance with an animated agent, will also vary across these age groups. For instance, one might hypothesize that an animated agent that is not designed with the needs of older adults in mind will distract from the successful completion of a task, rather than be of support towards that direction.

Other variables: Other user-related variables include attention capacity, background knowledge, culture, cognitive capability, and device experience.

Factor 2: Features of the Agent
Like users, animated agents can vary a wide variety across several features. These features include: Visual presence: Is there a need for interface agents to visually appear in interactive mobile guide interfaces? Due to scarcity of prior research in the area, it has not yet been investigated, if humanoid agents actually enhance mobile guide interfaces. It could be found, that voice output alone is sufficient for the interaction to take place successfully and for the task to be completed.

Modality of communication: Should the user be able to communicate, with the animated agent in a natural modality of communication, such as speech, or the use of alternative modalities, such as text input, or are menus with text phrases more appropriate for mobile environments? Given the complex technical challenge (such as, to distinguish the human voice and a car passing by) involved in creating an animated agent capable of accepting speech input, and the limited ability of users to enter text using a keyboard under mobile conditions, one can assume that the use of menus with text phrases, would be the most appropriate modality for communicating with the animated agent. However, if the technical problems can be solved, a full scale empirical evaluation can be made, to evaluate the three modalities of communication.

Other variables: Other agent-related variables are gender, amount of embodiment, ethnicity, and age, type of non-verbal cues, initiative, realism and personality.

Factor 3: Mobile Computing Environment:
The environment, in which the mobile guide operates, can of course vary in many ways. These include:

Type: We distinguish two types of mobile environments (of course other type of mobile environments are possible, such as that of a car) for which a mobile guide system and an animated agent must be tailored for: an indoor mobile environment (e.g., a museum or an art exhibition) and an outdoor mobile environment (e.g., a castle, ruins of an ancient temple, etc.). Because of the unique nature of each environment type, the potential effects of animated agents on the user, arc also distinctive. For example, in an outdoor environment the task of uncovering information about an attraction is by itself a less efficient task, than in an indoor environment, as the user must devote more mental and physical resources, to complete

the assigned task. For this reason, it cannot be assumed, that the existence of the animated agent, will result to the same effects on the user's perception and/or performance with the systems across the two environment types.

Characteristics: The characteristics of the environment in which the user is located (either indoors or outdoors), can vary in a variety of ways. These characteristics could affect the design of the mobile guide system and hence, the potential effects of the animated agent on the user. For instance, in some outdoor environments the dense built-up makes the use of location sensing equipment (e.g., a GPS device) a very ineffective method of navigation. The alternative methods that can be used (e.g., landmarks), will obviously have different kind of effects on the user (and thereby affect differently the user's perception and performance with the system).

Factor 4: Features of the Task
The tasks, in which the user is asked to perform with the aid of the agent, can also vary in many different ways. Some of these features are:

Navigational complexity: The complexity of the navigation task may be simple, in which the user has to find easy-to-find locations in an area. Alternatively, the user may be carrying out a more complex task, where he has to locate more hard-to-find sites in an area. The animated agent might have some impact (or no impact at all) on the perceived workload and performance of users, with the tasks of both complex levels.

Information personalization: An important feature of mobile applications is the ability to provide information, tailored to the location and profile of the user. Of course different levels of personalized information (e.g., according to the user's background knowledge), will have a different impact on the perception and performance with the animated agent. Although, more research is needed to define the appropriate levels and values for this variable, a systematic: exploration could be started, by considering the simple level of information difficulty, with values simple and technical information.

Future experiments, given the limited availability of technical and human resources, as well as time for the completion of the research, should address the following variables: user gender, user age, user perceptual capabilities, agent visual presence, task navigation complexness, and task information difficulty.

4 Experiment

4.1 Overview

An experiment has been conducted within the above research framework, to evaluate the impact of the mere presence of persona, on the accessibility and usability of a pedestrian mobile guide interface. The experiment was

conducted in the Monemvasia Castle, a common tourist destination in Greece, and manipulated the agent's visual presence (presence vs. absent) and order (present vs. absent and then vice versa) to observe any practise effects. The particular site was chosen because it has a large number of attractions, with a historical and cultural value and also junctions and decision points, in a relatively small area, without traffic, allowing a number of routes (both simple and complicated) to be tested. Usability and accessibility were evaluated, via both the performance and satisfaction dimensions. Based on findings from similar studies for stationary computing systems, we initially hypothesized, that the presence of the agent will have a positive effect on the user's subjective ratings of the systems, but his/her objective performance will likely not be affected by presence.

4.2 Participants

Nine participants were present in the field, to participate in this experiment and were randomly assigned to conditions. In order to avoid over-familiarity with the area of the Monemvasia castle, no participant was either a local resident or had visited the site before. All participants were native Creek speakers, but with good knowledge of the English language, and also had a variety of academic and mobile-computer backgrounds.

Fig.2 Screen Design of the Application in the present and absent condition

4.3 Procedure and Design

Participants were run individually using a tablet PC device with large display size and equipped with headphones. After a brief explanation (identical for all participants) about the purpose of the experiment from the experimenter, participants began the task, which was to navigate along a pre-selected route, visiting 3 locations in turn, and uncover information of interest about particular attractions. Participants were asked to perform this task, once using an interface with a visual agent, and the other using an interface with a non-visual agent. Since, we were interested in the effect of the mere presence of the agent; the difference between the two conditions concerned the presence of the agent only. Thus, all and only the information presented in the one condition were also presented in the other condition. No different tasks (i.e., different routes

and information contents) were employed because of the cost to run such an experiment in both human and technical resources.

After the participants, provided the system with sonic brief personal information (i.e., name and age), a computer agent (who supposedly has knowledge about the area) appeared, either as an animated character (right side of Fig. 2) or a disembodied voice (depending on the tested condition) (left side of Fig. 2) and proactively provided a brief overview of the history of the castle. The variable agent's visual presence, was manipulated within-subjects, while the variable order (present vs. absent and then vice versa) between-subjects. The visual agent was a 3D female (donated by DA Group), full-body, realistic (humanoid) character, capable of eye-blinking and movement of its mouth in synchronization with the synthesized voice. However, no gaze patterns, facial or body language, or any other form of nonverbal behaviours, were used to convey additional information. When the historical overview was completed, participants had the opportunity, to ask the agent certain questions, (from a list of pre-defined ones) about the functionality of the system, navigation, etc., and finally ask the agent to begin the tour.

Figure 7.3: Images from a segment of the first route

At the beginning of the route, the agent displayed the photograph of a landmark, and provided a brief verbal description, about how to get there and what to do next. For example, the speech instruction for the first photograph in Fig. 3 was, *"Passing through the main gate of the castle, you will find yourself in a dark portico, with a catacomb on your right, which was the post of the guards of the legendary gate. Continue your way, in the catacomb, and make the first turn on your right."* After tapping the next button, the participant was presented with a photograph of a new landmark and an audio instruction on how to get there. After visiting a certain number of landmarks, the participant arrived at the location, where he had to tap a symbol with the corresponding letter (different for each location) on the screen, in order to retrieve relevant information about the particular attraction (e.g. History/Architecture). The remaining of the photographs in Fig. 3, illustrate the landmarks that the user bad to follow in order to get to location A. After the presentation of information, the user could ask the agent questions from a list of static text phrases (different for each location the user was visiting) (shown in Fig. 2) below the character. When the presentation (and any possible questions) was complete, the user was able to move to the next location,

by tapping the button labelled "next" at the bottom left hand-side of the applications' screen. After the participants visited all the locations, they had to rate the systems on a 10th point scale, participate in a short interview, fill-in three questionnaires on their experiences with the systems, and finally take a, retention test on the information they heard during the tour.

4.4 Measures

4.4.1 Subjective Measures

Participant subjective impressions were measured by three questionnaires with items rated on 7-point Likert scales (1=strongly disagree, 7 = strongly agree). The first questionnaire examined the interaction between the participant and the systems. The second questionnaire assessed the participant's reactions to agents (visual arid non-visual). The third set of items referenced the subjective accessibility of the two systems. In addition, the experimenter asked participants a number of open-ended questions that provided participants an opportunity to give their impressions about the systems, as well as offer suggestions, about what should be improved in future versions.

Questions in the questionnaire for user subjective experience were formulated according to the ones from the literature (QUIS [4], IBM Computer Usability Satisfaction Questionnaire [5], SUMI [10], WAMMI [13], and covered aspects of satisfaction around the following three indices: ease of use (e.g., "I think the application was difficult to use), efficiency (e.g., "The application is too slow"), likeability (e.g.," Overall I am quite satisfied with the application") and user feelings (e.g., "frustrating" and "confusing"). In the same way, the agent specific: questions were based on dimensions of the user's subjective experience that are commonly measured in the literature [5]. This questionnaire was divided into two parts (Q-A and Q-l3). The first part included questions that concerned all agents, while the second part concerning only the visual agent. Both parts addressed a number of qualities of the agents, like "personality", "helpfulness", "intelligence", etc. The accessibility questionnaire was based on the Simplex II model [2], and included items like, complexity of the tasks (i.e., information and navigation), learnability of the tasks, etc. However, no questions related to the effectiveness of the animated agent, in guiding the user's attention on the environment (see embodied tenet of distributed cognition), were included in this questionnaire.

4.4.2 Objective Measures

Towards the more objective end, the performance of participants on the tasks, in terms of time, navigational errors, and retention performance was measured. With regards to time and navigational errors, a direct observation for note taking method, was employed. In particular, on the routes, the experimenter walked a few steps behind the participant, in order not to influence, navigation decisions and information presentation. He made

written observations on general behaviour, as well as, provided help when participants faced some sort of problem. The time taken and the number of times that participants got lost on the chosen route, by using each of the systems were also measured by the experimenter. A participant was defined as lost, if the experimenter had to intervene, to get him/her back onto the route. Retention performance was measured, by the participants' answers in a short retention test.

5 Results

5.1 Subjective impressions

For the analysis of the usability questionnaire, a 2 X 2 ANOVA was conducted on each of the questionnaire items, taking visual presence and order, as independent variables and ratings, as the dependent variable. We found significant main effects and interactions on the entertainment of information presented by the systems; likeability of the dialogues and on whether participants get what they expect, when clicking on objects of the application.

On the entertainment, of information presented by the systems, there was a significant effect of present, vs. absent; ($F(1, 14) = 5.303$, $p < 0.05$) and a significant interaction between present vs. absent and P-A vs. A-P ($F(1, 14) = 5.303$, $p < 0.05$). A one-way ANOVA test for visual presence across the order conditions, showed that participants in the A-P condition, viewed the system with the visual agent significantly less entertaining, than the system with the non-visual agent ($F(1,4)$ 12.500, $p < 0.05$). Additionally, ANOVA tests for visual presence between the order conditions, nearly reached statistical significance for present in the A-P order ($F(1,7) = 5.33$, $p = 0.054$), suggesting that participants in the particular condition, viewed the information presented by the system with the visual agent as less entertaining, than participants in the P-A condition. No other ANOVA comparisons reached significance level.

On the likeability of the dialogues, there was a highly significant main effect of P-A vs. A-P; $F(1,14)$; 10.667, $p < 0.01$). A one-way ANOVA test across the order conditions failed to reach statistical significance, either for present or absent. No effect, for visual presence, between the order conditions, was found either. This suggests that participants across the order conditions, perceived both agents significantly differently, but their perceptions of each agent type, did not differ between or across the order conditions. A close inspection of the data, reveals that participants in the A-P condition, gave more negative ratings (mean = 3.66) for both agents than did participants in the P-A condition (mean = 6.1).

On whether participants get what they expect, when clicking on objects of the application, there was a highly significant main effect of P-A vs. A-P ($F(1,14) = 11.339$, $p < 0.01$). One-way ANOVA tests revealed that participants in the A-P condition viewed the system with the non-visual agent, as

significantly less responsive, when clicking on its objects, than participants in the P-A condition.

For the analysis of part A of the assistant-specific questionnaire, we performed a 2 X 2 ANOVA on each of the items, taking visual presence and order, as independent variables and ratings, as the depended variable. We found no significant effects, either for agent type or order, for any of the questions. For the items in part B of the questionnaire, we performed a one-way ANOVA, using the same variables. We found significant effects on the helpfulness of the visual agent, in the navigation task and the helpfulness of the visual agent, in the information task.

On the helpfulness of the visual agent on the navigation task, there was a significant, main effect of P-A vs. A-P ($F(1,7) = 16.608$, $p < 0.01$). Participants in the A-P condition (mean = 3) viewed the visual agent significantly less helpful in the navigation task, than did participants in the P-A (mean = 6.6) condition.

The observed difference for P-A vs. A-P, was even larger for the helpfulness of the visual agent in the information task ($F (1,7) = 112.000$, $p < 0.01$). Participants in the A-P condition (mean = 2.6), considered the visual agent, significantly less helpful in the information task than participants in the P-A (mean present = 6.6) condition.

Lastly, with regard to the accessibility questionnaire, we performed a 2 X 2 ANOVA on each of the questionnaire items, taking visual presence and order, as the independent variables and ratings, as the dependent. We found significant main effects on the audibility of the dialogues, and on whether the structure of the tour was presented well enough.

On the audibility of the dialogues, there was a significant, main effect of P-A vs. A-P ($F (1,14) = 5.744$, $p< 0.05$). A one-way ANOVA test showed, that participants in the A-P condition viewed the dialogues with the visual-agent as significantly more audible, than the participants in the P-A condition ($F(1,7) = 14.913$, $p < 0.01$). No other ANOVA comparisons reached statistical significance at .05 level.

On whether the structure of the tour was presented well enough, there was a significant main effect, of PA vs. A-P ($F (1, 14) = 9.692$, $p < 0.01$). A one-way ANOVA was calculated on participants' ratings of agent types across the order conditions, but failed to reach statistical significance for either present or absent. There was no effect for visual presence between the order conditions either. This shows that participants across the two order conditions, perceived both agent types significantly differently, but their impressions of each agent type did not differ between or across the order conditions. A close examination of the collected data, shows that participants in the A-P condition, gave on average more positive ratings (mean = 2.833) for both agents, than did participants in the P-A condition (mean = 1.333).

5.2 Task Performance

As mentioned earlier, participants' performance was measured in terms of the total time to complete a tour with the system, the number of errors conducted and their performance in the short retention test. For all analysis, we performed a 2x2 ANOVA, taking visual presence and order, as independent variables and time (in seconds) and scores in the test, as dependent variables. With respect to the number of times that participants got lost, we did not attempt any statistical analysis on the collected data, as the numbers were too small. We found significant interactions for visual presence, and the time taken to complete a tour, and no significant effects or interactions for visual presence or order on the participants' retention performance.

On time there was a significant, interaction between present vs. absent, and P-A vs. A-P ($F(1, 14) = 7.956$, $p < 0.05$). This interaction suggests that participants' time performance was different with agents of the same type, across the two order conditions. In order to investigate this further, we ran a series of one-way ANOVA tests on the collected data. For the present condition, the ANOVA test failed to reach statistical significance, while for absent the ANOVA test, showed a significant effect of order ($F(1, 7) = -17.596$, $p < 0.05$), with participants in the A-P condition spending significantly more time with the non-visual agent, than participants in the P-A condition. However, a one-way ANOVA, for absent vs. present across the two order conditions, did not show any significant results, thus revealing that this difference was a chance result. Then, an ANOVA test for visual presence between the order conditions, showed a significant effect, for the A-P ($F(1, 4) = 22.154$, $p < 0.05$), as well as for the P-A ($F(1, 10) = 6.865$, $p < 0.05$) condition.

As regards to the ratings that participants gave for each of the systems, we conducted a 2 X 2 ANOVA, taking visual presence and order as independent variables, and the mean ratings as the dependent variable. We found no significant main effects for visual presence or order. No significant interactions were found either:

5.3 Interview Feedback

The analysis of the responses to the interview questions, confirmed the quantitative findings, and provided additional insights about the animated agent and the applications. Two of the participants in the assigned conditions, considered the visual agent as a distraction in both the navigation and information tasks, but stated that it was less distracting in the navigation and more in the information task. Five of the participants found the visual agent to be a distraction in the information task, whereas only two participants considered the visual agent useful in the navigation task. Two participants felt that the visual agent was a distraction in the information task, but functionally neutral (neither a helper nor a distraction) in the navigation task.

Virtually all participants found the speech output from both systems of poor quality. Participants expressed a high discomfort with this feature, as it was very difficult to understand the information provided. This makes sense, because the agent was providing subjects with information inertly using speech, and sonic slight body movement. In a full multimodal system, where the agent would have the ability, to generate the proper nonverbal behaviours, to supplement or complement the information provided, the effect of the speech distortion on the user would probably have been limited. Six participants in the chosen conditions expressed negative views about the amount and the accuracy/relevancy of the information presented by the systems. Some of their comments included: "I think that the application presents too much information that the user would probably not be interested in", "I found that there were some historical inaccuracies to the information presented by the system". It appears that those participants managed to overcome the speech distortion, and focus their attention on the understanding of the presented information.

As for improvements, participants suggested a variety of ideas. Three participants explicitly recommended improvements in the question-answer dialogue process of the systems. They suggested improving the dialogue turn-taking, aspects of the systems' feedback, the quality and quantity of the information provided by the questions and others. Their comments included: "...the system should provide more "in-depth" answers to the questions selected by the user."; "...I would like to be given the option of making my own questions and not selecting from a. list of predefined ones". Then, all but two participants actively requested, embodying the visual agent with appropriate non-verbal behaviours. Four participants suggested improving the appearance of the visual agent. One participant suggested, that the visual agent should remain visible in the navigation task, but in the information task, it, should either become transparent or reduce in size, and move to a, corner of the screen. Finally, two participants suggested, replacing the electronic voice with the voice of a real human.

6 Discussion

Overall, this experiment examined the possible effects of the mere presence of an animated agent, on the GUI of a mobile guide system. Quantitative statistical analysis showed, that the present vs. absent and P-A vs. AP manipulation, had a significant influence on a number of questionnaire items (with differences mainly in the order of the presentation), as well as the participants' time performance. Below, we separately discuss these results, along with some interesting hypotheses that they suggested, in more detail.

6.1 Subjective Assessment

Although the results reported above, should be interpreted with caution, they do make sense. With respect, to the usability questionnaire, and on the effect on the entertainment value, the A-P task required participants to experience the visual agent after the non-visual agent, thus adding a further

object to attend to, in the already known interface. Because of this, participants may have while for absent the ANOVA test, showed a significant effect of order ($F(1, 7) = -17.596$, $p < 0.05$), with participants in the A-P condition spending significantly more time with the non-visual agent, than participants in the P-A condition. However, a one-way ANOVA, for absent vs. present across the two order conditions, did not show any significant results, thus revealing that this difference was a chance result. Then, an ANOVA test for visual presence between the order conditions, showed a significant effect, for the A-P ($F (1, 4) = 22.154$, $p < 0.05$), as well as for the P-A ($F(1, 10) = 6.865$, $p < 0.05$) condition.

As regards to the ratings that participants gave for each of the systems, we conducted a 2 X 2 ANOVA, taking visual presence and order as independent variables, and the mean ratings as the dependent variable. We found no significant main effects for visual presence or order. No significant interactions were found either:

5.3 Interview Feedback

The analysis of the responses to the interview questions, confirmed the quantitative findings, and provided additional insights about the animated agent and the applications. Two of the participants in the assigned conditions, considered the visual agent as a distraction in both the navigation and information tasks, but stated that it was less distracting in the navigation and more in the information task. Five of the participants found the visual agent to be a distraction in the information task, whereas only two participants considered the visual agent useful in the navigation task. Two participants felt that the visual agent was a distraction in the information task, but functionally neutral (neither a helper nor a distraction) in the navigation task.

Virtually all participants found the speech output from both systems of poor quality. Participants expressed a high discomfort with this feature, as it was very difficult to understand the information provided. This makes sense, because the agent was providing subjects with information inertly using speech, and sonic slight body movement. In a full multimodal system, where the agent would have the ability, to generate the proper nonverbal behaviours, to supplement or complement the information provided, the effect of the speech distortion on the user would probably have been limited.

Six participants in the chosen conditions expressed negative views about the amount and the accuracy/relevancy of the information presented by the systems. Some of their comments included: “I think that the application presents too much information that the user would probably not be interested in", "I found that there were some historical inaccuracies to the information presented by the system". It appears that those participants managed to overcome the speech distortion, and focus their attention on the understanding of the presented information.

As for improvements, participants suggested a variety of ideas. Three participants explicitly recommended improvements in the question-answer dialogue process of the systems. They suggested improving the dialogue turn-taking, aspects of the systems' feedback, the quality and quantity of the information provided by the questions and others. Their comments included: "...the system should provide more "in-depth" answers to the questions selected by the user."; "...I would like to be given the option of making my own questions and not selecting from a. list of predefined ones". Then, all but two participants actively requested, embodying the visual agent with appropriate non-verbal behaviours. Four participants suggested improving the appearance of the visual agent. One participant suggested, that the visual agent should remain visible in the navigation task, but in the information task, it, should either become transparent or reduce in size, and move to a, corner of the screen. Finally, two participants suggested, replacing the electronic voice with the voice of a real human.

6 Discussion

Overall, this experiment examined the possible effects of the mere presence of an animated agent, on the GUI of a mobile guide system. Quantitative statistical analysis showed, that the present vs. absent and P-A vs. AP manipulation, had a significant influence on a number of questionnaire items (with differences mainly for order of the presentation), as well as the participants' time performance. Below, we separately discuss these results, along with some interesting hypotheses that they suggested, in more detail.

6.1 Subjective Assessment

Although the results reported above, should be interpreted with caution, they do make sense. With respect, to the usability questionnaire, and on the effect on the entertainment value, the A-P task required participants to experience the visual agent after the non-visual agent, thus adding a further object to attend to, in the already known interface. Because of this, participants may have certainly become more attentive to the information itself, thus noticing more problems with it (like, its low entertainment value), than participants in the P-A condition without the animated agent. Then, the group difference for the likeability of the dialogues could certainly have been due to the strong negative reactions, towards the general dialogue features of the two systems. The negative reactions towards various aspects of this feature, as noted in the interviews, seem to support this view. Lastly, even though the results for whether objects react as participants expect when they click on them, show a significant difference for absent, it is evident from the collected data, that participants in the AP condition had more problems interacting with objects on both applications (mean=4.167), than participants in the P-A condition (mean=6.333). This tendency was also recorded in the field, where participants in the A-P condition reported problems with the use of the device stylus, in order to interact with the applications. No such observations were made for the participants in the P-A condition.

As regards to the assistant-specific questionnaire, it is evident that practice provoked negative reactions, towards the helpfulness of the visual agent in the navigation, and information tasks. It appears that once the participants become familiar with the functionalities and features of the mobile guide system, they could express Letter and stronger opinions, on the helpfulness of the visual agent in the assigned tasks. The data retrieved in the interviews from the participants in the A-P condition seem to reiterate this point.

Lastly, the results from the accessibility questionnaire reveal two important implications for the design and implementation of our future systems and experiments. For the audibility of the dialogues, the stronger explanation is, that when we ran participants in the P-A condition with the visual agent, external environmental factors (e.g., noise) limited their ability to hear the dialogues with the visual agent efficiently. However, it can be argued that if the visual agent would have provided an additional channel (e.g., through nonverbal cues) for participants to perceive the dialogues with the system, this environmental effect would probably have been minimal. On the other hand, however, the group difference in whether the structure of the tour was presented well reveals a factor that can affect, this multimodal relationship, i.e., the individual perceptual (a variable already predicted in our research framework). For example, if the P-A group was consisted predominately from visual users, and the A-P group from auditory users, then the structure of the tour would certainly have seemed as it was presented less well to the first group of users than in the second one.

From the above discussion it is clear, that although, the assessment of the user's subjective experience of the systems, has contributed a number of important findings/ hypotheses to our follow-on experiments, the key findings (always with caution) are: 1) the presence of an animated agent, may affect the user's view of the entertainment value of the systems, with this effect interacting with the order of presentation, and 2) that practice resulted from the presentation order manipulation, may play a role in the perception of the agents.

6.2 Objective Performance

The results reported for time, clearly show a strong practice effect on the participants' performance, with the suggestion that this effect might be modified by the present or absent conditions. This suggests an important implication for our future experiments, and an interesting hypothesis. With regards to the first, this finding suggests that, we will need to be aware of the practice effects in future experiments, either as modified by agent effects or independent of agent effects. With regards to the second, we strongly believe that this practice effect was due to a strong familiarity with the question-answering dialogues of the two systems. Because, these were the same between the agent conditions, it is possible that participants asked the systems less questions in their second attempt, and thereby, spent less time, than in the first. A close inspection of the collected data, shows

that this difference, was larger with the non-visual agent (mean = 1283.0(1) than with the visual-agent (mean = 1398.22). This suggests our first hypothesis: *Hypothesis 1: The presence of the animated agent increases the time taken to complete a tour, for instance,* by *stimulating participants to explore (through question-asking) the information available about a, location in more-depth.*

With regards to the navigational errors, although we didn't perform any statistical analysis, a careful examination of the collected data, can reveal useful insights for future studies. In particular, participants got lost (where lost is defined above) with the visual agent less often (mean = 0.2 times), than with the non-visual agent (mean = 0.4 times). These data suggest our second hypothesis. *Hypothesis 2: An animated agent could enhance the user's navigational ability, for instance* by *helping the user understand the underlying structure, of the physical space* and *hence, allow him to better navigate himself.*

In terms of retention performance, we noted the following: although, the information contents of the systems between the agent conditions were the same, participants on average, did better with the non-visual agent (mean = 5.06) than with the visual agent (mean 5.06). This suggests our third hypothesis. *Hypothesis 3: An animated agent decreases the user's retention performance, for instance, by distracting participants away from the presentation of information about a location.* However, based on the comments of participants in the interviews, we can derive a contradictory hypothesis. *Hypothesis 4.- The presence of the animated agent increases retention performance in the user, for instance, because, given an agent, capable of generating appropriate non-verbal behaviours to accompany the linguistic information, it renders the interaction with the system smoother, thus potentially supporting greater retainability.*

Finally, in relation to the ratings that participants gave for the two systems, it is interesting that the participants gave overall better ratings for the system with the non-visual agent (mean = 8.12) than for the system with the visual agent (mean = 7.67). More than half of the participants indicated, that they liked the non-visual agent more than the visual agent.

From this discussion, it is clear that the visual agent did not foster any kind of significant influence on the participant's performance. However, one would expect that based on the influence of visual presence on the users' subjective views, there should have been sonic kind of impact on their objective task performance as well. One possible explanation is that because the animated agent was not capable of generating non-verbal output, its mere presence on the interface, from the one hand, fostered participants subjective views of its unhelpfulness in the tasks, and from the other hand, failed to construct some kind of mental model (of both locations and routes) that could have led to enhanced or worse performances.

7 Conclusion

Whilst a large amount of empirical work, has contributed to the evaluation of animated agents for stationary computer systems, when it conies to the user of mobile applications, there is currently a scarcity of previous research in the area. In addition, the very few existing studies, lack of a solid framework for systematically examining the impact of anthropomorphic animated agents, on the user's performance and subjective experience of mobile applications. We developed a theoretically well-supported five factor approach, for studying animated agents. We performed an experiment, within this framework that suggested, that the presence of an animated agent may affect the user's subjective experience of the mobile guide system with this effect interacting with the order of presentation, and that practice resulted from the presentation order manipulation may play a role in the perception of agents. Additionally, performance data revealed that in future experiments, we will need to be aware of practice effects, either as modified by agent effects or independent of agent effects, and a number of hypotheses on what to expect from these works. We plan to use our framework, to guide three additional empirical studies, in which the first will be the continuation of this experiment, and the other two will examine separately, the problems of information provision and navigation, in the same outdoor mobile conditions.

References

1. Adams, R. and Langdon, P. (2003): "SIMPLEX: a simple user check-model for Inclusive Design." In: Universal Access in IICI: Inclusive Design in the Information Society. Stephanidis, C. (Ed.). 4, 13-17. Mahwah: NI: Lawrence Erlbaum Associates.
2. Adams, R. (2007): "Decision and stress: cognition and e-accessibility in the information Workplace" In: Universal Access in the Information Society (UAIS), Volume 5, Number .4, April 2007, pp. 363-379(17)
3. Bider, D., Haussler„I., Kruger, S. Nlinker, W. (20(12): "The SmartKom Mobile MultiModal Dialogue System". In: E. Andre, L. Dybkjaer, P. Heisterkamp, W. Mink,'r (Eds.), Extended Abstracts of the ISCA Tutorial and Research Workshop on MultiModal Dialogue in Mobile Environments (p. 35-38). Kloster Irsee, Germany: University of Southern Denmark.
4. B. Harper and K.Norman, Improving User Satisfaction: The Questionnaire for User Interaction Satisfaction Version 5.5, Proceedings of the Annual Mid-Atlantic Human Factors Conference, Virginia Beach, VA, 1993, pp. 224-228
5. B.Shneiderman, Designing the User Interface. Strategies for Effective Human-Computer Interaction, 3rd Edition, Addison-Wesley, Reading, MA, 1998
6. Cowell, A. J., Tana.sse, T. E., Stanney, K. M. (2003): "Using anthropomorphic embodied conversational agents in mobile guides and information appliances". 5th International Symposium on Human Computer Interaction with Mobile Devices and Services, Mobile HCI '03, Udine, Italy, September 8-11, 2003.
7. Catrambone, R., Stasko, 1, X iao„.1. (2002): "Anthropomorphic agents as a

user interface paradigm: Experimental findings
and a framework for research", Proceedings of the 24th Annual Conference of the Cognitive Science Society, Fairfax, VA, August 2002, 166-171

8. Dehn, D. NI., Van Mulken, S. (2000): "The impact of animated agents: a review of empirical research." Int. Human Computer Studies 52. 1-22.
9. Hollan, .1., Hutchins, E., Kirsh, 1). (2000): "Distributed Cognition: Toward a New Foundation for Human Computer Interaction Research". ACM Trans. on Computer-Human Interaction. 7 (2), 174-196
10. Hollan, .1., Hutchins, E., Kirsh, 1). (2000): "Distributed Cognition: Toward a New Foundation for Human Computer Interaction Research". ACM Trans. On
11. HFRG, Human Factors Research Group, Software Usability Measurement Inventory, SUMI, 1993,http:www.ucc.ie/hfrg/questionaires/sumi
12. Lai,J., Wood D., Considine, M.(2000): “The effect of task conditions on the comprehensibility of synthetic speech”, In: Proceedings of the ACM CHI 2000, 321-328
13. McBreen, H., Jack, M., (2001): “Evaluating Humanoid Synthetic Agents in E-Retail Applications”, IEEE Transactions on Systems, Man and Cybernetics, Part A, Systems and Humans, vol.31, no.5,394-405
14. WAMMI consortium-Web Usability Questionaite, 2002, http://www.wammi.com
15. Xiao j., Stasko, J., Catrambone, R. (2004): “An Empirical Study on the Effects of Agent Competence on User Performance and Perception”. Proceedings of AAMAS ’04,NewYork, NY July 2004, 178-185.

APPENDIX D:

This appendix contains several tables. The tables contain data from the empirical studies that were conducted to evaluate three of the six prototype mobile tour guide systems. Full details of the analysis conducted on the data can be found in Chapter 7. "Present" (P) indicates that the ECA is present. "Absent" (A) indicates that the ECA is absent.

Experiment One:

Participants:

Order	Name	Age/Gender	Profession	Mobile device user
P/A	001	45/Female	Social worker	2G phone
P/A	002	33/Male	Accountant	2G phone
P/A	003	33/Male	Optician	3G phone/laptop
P/A	004	40/Female	Manager	2G phone
P/A	005	34/Female	Reporter	2G phone
P/A	006	53/Female	Cook	2G phone
P/A	007	33/Male	Naval Engineer	2G phone/laptop
P/A	008	29/Male	Web designer	3G phone
P/A	009	32/Female	Shop owner	2G phone
A/P	010	22/Male	Student	3G phone
A/P	011	19/Male	Student	3G phone
A/P	012	30/Female	Shop assistant	2G phone
A/P	013	33/Female	Teacher	2G phone
A/P	014	25/Male	Web Designer/Game Tester	3G phone
A/P	015	30/Male	HR Assistant/Game Tester	3G phone

A/P	016	32/Female	Commercial Manager	3G phone
A/P	017	34/Male	Software engineer	3G phone
A/P	018	31/Male	Student	3G phone

Table D.1.1: Participants in experiment one

Objective Assessment:

Present (n=18)	Absent (n=18)	Order
1653	979	P/A
1553	978	P/A
1891	1213	P/A
1476	975	P/A
1810	1014	P/A
2222	2015	P/A
1238	1301	P/A
1927	1384	P/A
1913	1038	P/A
1123	1638	A/P
1158	1438	A/P
1176	1957	A/P
1055	1437	A/P
1185	1659	A/P
926	1839	A/P
1022	1324	A/P
1064	1629	A/P
1374	1214	A/P
1431.4	**1390.1**	**Mean**

Table D.1.2: Time taken (in seconds) to complete the tour in experiment one

Present (n=18)	Absent (n=18)	Order
2	2	P/A
2	4	P/A
1	2	P/A
3	3	P/A
1	2	P/A
0	2	P/A
0	3	P/A
0	1	P/A
0	3	P/A
1	1	A/P
5	1	A/P
4	1	A/P
4	1	A/P
3	2	A/P
0	0	A/P
0	1	A/P
2	0	A/P
0	1	A/P
1.55	**1.66**	**Mean**

Table D.1.3: Frequency of getting lost in experiment one

Present (n=18)	Absent (n=18)	Order
5	6	P/A
6	7	P/A
7	6	P/A
5	7	P/A
12	8	P/A
25	43	P/A
6	14	P/A
13	9	P/A
15	7	P/A
7	6	A/P
7	8	A/P
4	5	A/P
9	8	A/P
6	6	A/P
4	7	A/P
7	9	A/P
7	9	A/P
13	8	A/P
8.7	**9.6**	**Mean**

Table D.1.4: Total questions asked in experiment one

Present (%) (n=18)	Absent (%) (n=18)	Order
38	33	P/A
31	0	P/A
31	22	P/A
73	41	P/A
19	4	P/A
15	22	P/A
33	22	P/A
54	41	P/A
15	0	P/A
4	35	A/P
15	23	A/P
22	19	A/P
56	69	A/P
4	27	A/P
4	46	A/P
33	50	A/P
44	54	A/P
2	27	A/P
27.3	**29.7**	**Mean**

Table D.1.5: Participants' retention scores in experiment one

Subjective Assessment:

Order	**Present (Y/N)** **(n = 18)**	**Absent (Y/N)** **(n = 18)**
P/A	5/4	7/2
P/A	7/2	7/2
P/A	6/3	7/2
P/A	7/2	6/3
P/A	6/3	8/1
P/A	6/3	7/2
P/A	6/3	8/1
P/A	4/5	4/5
P/A	7/2	7/2
A/P	6/3	5/4
A/P	6/3	6/3
A/P	5/4	6/3
A/P	8/1	6/3
A/P	7/2	2/7
A/P	8/1	8/1
A/P	6/3	7/2
A/P	6/3	6/3
A/P	8/1	7/2
Total:	**114/48**	**114/48**

Table D.1.6: The object recognition (Yes/No) results in experiment one

Questions	P/A (n=18)		A/P (n=18)		AVG
	P	A	A	P	P/A
1) The navigation and information task are too complex	2	2	2	3	3/2
2) The navigation and information task are difficult to learn	3	2	2	3	3/2
3) The process of navigation and information extraction from the systems is difficult to learn	3	2	2	2	3/2
4) The screens are not consistent with the navigation instructions	2	3	2	3	3/3
5) The screens are not consistent with the information given about a location	2	3	2	3	3/3
6) The completion of the information and navigation tasks require too much self-organization	3	4	3	3	3/4
7) It does not give adequate visual input	2	2	1	2	2/2
8) It does not give adequate auditory input	2	1	2	3	3/2
9) The methods of information presentation (i.e., voice, images, gestures and face expressions) are many and confuse me. I would like a simpler system (e.g., with voice or text)	2	2	3	4	3/3
10) It is difficult to understand the dialogues used by the system	3	2	2	3	3/2
11) The information presented by the system is poorly presented (too brief or too long)	2	2	2	2	2/2
12) The output the system (i.e., audio, gestures, face expressions and images) is poorly timed	1	1	2	3	2/2
13) The output of the system (i.e., audio, gestures, face expressions and images), is unclear	2	1	2	3	3/2
14) The output of the system (i.e., audio, gestures, face expressions and images), is not relevant with the topic at hand	1	1	2	2	2/2
15) I have to hold too much information in mind to navigate in the castle.	3	3	2	3	3/3

16) After visiting a location, I find it hard to remember the information that was presented about this location	4	4	3	4	4/4
17) I have to hold too much information in mind when using the system	3	3	4	4	4/4
18) I have to think carefully before responding to the information presented by the system	3	3	3	3	3/3
19) The system should prompt me to pay attention to a presentation about a location	2	3	4	4	3/4
20) The system should automatically respond (e.g., with the pause of a presentation) when I am confused or overloaded with information.	5	6	4	5	5/5
21) The system is too frustrating to use	1	2	2	2	2/2
22) The system is too annoying to use	2	2	3	2	2/3
23) The design of the system is not serious enough	1	1	2	2	2/2
24) The system is fun to use	6	6	5	5	6/6
25) The design of the system makes it difficult to learn what I need to learn to use it properly	2	2	3	3	3/3
26) It's hard to learn the information presented about a location or how to navigate in the castle	2	2	2	3	3/2
27) The information presented by the system should relate better to what I already know	3	3	2	4	4/3
28) The information scenarios should related better to my personal interests	3	3	3	3	3/3
29) I find it difficult to construct a mental map of the route in the castle as it is presented by the system	3	2	4	3	3/3
30) The structure of the information presented about a location is difficult to follow	2	2	3	4	3/3
31) The modalities used by the system (i.e., voice, images, gestures and face expressions) prevent me from building a clear "mental picture" of the information presented about a location	3	3	3	3	3/3

32) The modalities used by the system (i.e., audio, images, gestures and face expressions) prevent me from building a "mental map" of our route in the castle.	3	3	2	3	3/3
33) The system should make allowances for my response errors (for example during navigation)	5	5	5	5	5/5
34) The system does not provide me with sufficient information to respond with appropriate reactions to its requests (where to go or what to do next)	2	2	3	3	3/3
35) The system requires me to make unreasonable responses (e.g., navigate hard-to-walk routes)	2	2	2	2	2/2
36) I make a lot of response errors with the system (i.e., wrong navigation decisions or wrongly retained information)	2	3	2	2	2/3
37) The system requires me to find landmarks and/or locations that are too difficult to find	2	2	2	3	3/2
38) I always have to seek the experimenter's help to proceed from location to location.	3	2	2	2	3/2
39) The system gives me no support to learn the information it presents about a location.	1	1	2	3	2/2
40) I never know the correct navigational instructions in order to get to my destination	1	2	2	2	2/2

Table D.1.7: Mean responses to the cognitive accessibility questionnaire

Questions	**P/A (n=18)**		**A/P (n=18)**		**AVG**
	P	**A**	**A**	**P**	**P/A**
1) I think the system is difficult to use	2	2	1	2	2/2
2) The structure of the system makes it difficult to navigate it	2	2	2	2	2/2
3) The dialogue window makes it difficult to ask the system	3	2	1	3	3/2
4) The system uses terms understandable and familiar to me	6	6	7	6	6/7

5) The system has too many choices	4	5	1	4	4/3
6) It is difficult to tab on objects of the system	3	2	1	4	4/2
7) The system is too slow	3	2	1	2	3/2
8) I get what I expect when I click on objects of the system	6	6	7	5	6/7
9) I need time to familiarize myself with the system before the tour begins	4	3	1	3	4/2
10) I had to pay too much attention to the system to complete the tasks	4	3	3	3	4/3
11) I find this system useful for navigating and extracting information about the castle	6	6	7	6	6/7
12) Compared to what I expected, the tasks did go really quickly	6	5	5	5	6/5
13) Using the system was an engaging experience	6	6	7	5	6/7
14) I found the information presented by the systems entertaining	5	6	7	4	5/7
15) I thought that my conversation with the system was unnatural	3	2	1	3	3/2
16) The system is an excellent idea to make tourism an interactive experience	6	6	7	6	6/7
17) The system is innovative	6	6	7	5	6/7
18) Overall I am satisfied with the system	6	6	7	5	6/7
19) Annoying	1	1	1	2	2/1
20) Confusing	2	3	2	2	2/3
21) Frustrating	1	1	1	2	2/1
22) Interesting	6	6	7	6	6/7
23) Intelligent	7	6	7	6	7/7
24) Refreshing	6	6	7	4	5/7
25) Tiresome	2	3	1	3	3/2

26) unpleasant	1	2	1	2	2/2

Table D.1.8: Mean responses to the usability questionnaire

Questions	P/A (n=18)		A/P (n=18)		AVG
	P	A	A	P	P/A
1) The virtual guide distracted me from the tasks	3	3	3	4	4/3
2) The virtual guide was friendly	6	7	6	6	6/7
3) The virtual guide was annoying	1	2	2	3	3/2
4) The virtual guide was intelligent	5	6	5	5	5/6
5) The virtual guide was competent	5	6	5	5	5/6
6) The virtual guide was emotionless	3	2	2	3	3/2
7) The virtual guide was demanding	2	2	2	2	2/2
8) The virtual guide was polite	6	7	6	6	6/7
9) I liked the voice of the virtual guide	6	6	6	5	5/6
10) The voice of the virtual guide was not clear enough	3	2	3	5	5/3
11) The voice of the virtual guide was not appropriate for the system	2	2	2	2	3/2
12) I would prefer a more natural voice for the virtual guide	5	5	4	4	4/5
13) I like the appearance of the virtual guide	6	N/A	N/A	5	5
14) I would prefer a more realistic virtual guide	4	N/A	N/A	5	5
15) The appearance of the virtual guide distracted me from the tasks I had to complete	2	N/A	N/A	3	3
16) The appearance of the virtual guide is not appropriate for this system	2	N/A	N/A	3	3
17) The gender of the virtual guide is not appropriate for this system	6	N/A	N/A	5	5

18) The lip-synchronization of the virtual guide distracted from the tasks	2	N/A	N/A	4	4
19) A virtual guide capable of face-detection and generation of appropriate responses is the minimum interactive feature such a system should have	4	N/A	N/A	4	4
20) The virtual guide help me in the disambiguation of the information during a presentation about a location	5	5	5	5	5/5
21) The virtual guide should help me in erroneous situations (e.g., when I am lost in a route)	6	6	6	6	6/6
22) It would have been impossible to complete the tasks without the help of the virtual guide (compared to a guide book)	4	5	4	4	4/5
23) The body language (gestures and face expressions) made the virtual guide to look "natural"	5	N/A	N/A	4	4
24) The guide's body language made her look complete	4	N/A	N/A	4	4
25) The guide's body language made her look non-friendly	3	N/A	N/A	3	3
26) I liked the guide's body language	5	N/A	N/A	5	5
27) The guide's body language looks realistic	4	N/A	N/A	4	4
28) The guide's body language looks excessive	2	N/A	N/A	3	3
29) The guide's body language was not relevant with the presented information	2	N/A	N/A	3	3
30) The guide's body language was not correctly synchronised with her speech	1	N/A	N/A	2	2

Table D.1.9: Mean responses to the ECA-specific questionnaire

Interviews - Order 1 (Present vs. Absent):

Question 1: **Do you have any comments about the systems?**

The avatar could have been more natural and approachable. I think some of the local attractions of the castle are not exactly brought out with the system. The avatar could have included more humor in the descriptions, and those should be

U1: combined with visualizations (or animations) of the various historical events. Some of the questions were tiresome (mainly due to lack of focus). Finally, I think you need to include one more scenario that of the local tradition. There are local traditions and values, beautiful scenery in the castle, and local dances that will certainly attract the attention of the visitor.

U2: I think both systems are quite good

U3
- Choice between gender (male and female)
- No parameterize (pre-choice of guide – made by the manufacturer)
- Problems with the Panoramic Applications (confused with the E button – would prefer arrows)
- Prefer a system that adapts to the user (go wherever you want instead of pre-made routes)

U4:
- Enjoyable, Useful
- Interesting
- Innovative

U5: No

U6:
- Must be easier to ask questions. Free input with NLP
- Less dates.
- Less speed in the description. All a user is left is with the impression that the locations in the castle have important historical value... it's impossible though to remember anything else.
- Keep the feeling I have from the whole thing

U7:
- Voice is not clear
- In the second system things were better as the guide was not present. The character attracted my attention.
- Had some problems with the navigational instructions. They were not clear enough.
- Had some problems in the second system at the "Elkomenos" church (specifically the snakes above the temple)
- The second system showed three churches – she confused the information presented for each church
- The second system used terms not familiar to me. I completely lost the presented information because of that.

U8: It was a bit confusing (The navigation instructions of the second part)

U9:
- User should be allowed to interrupt the character while she speaks (Hardware problem)
- Scrolling of the Dialogue window Should be there

Question 2: **How do you think I can improve the design of the system?**

U1: See above

U2: I don't think any of the systems can be improved in any way

U3:
- less Information (Simple Information)
- Less speed (Location C is the worst of all)

U4:
- Not so many dates in the history scenario
- Zoom at the points where the character talks about
- The guide should talk a bit slower

U5:
- Louder voice
- Larger screen

U6:
- Very cold. How is it possible to make it more human?
- The guide should make her to ask question and not pass them like that (typically).
- If you have a human you can also ask more stupid questions.

U7:
- The guide attracted my attention too much. Not the appearance but the gestures of the guide. When the information is not important to me I mostly look at the guide.
- Above the Buttons I should put short textural description (e.g., temple, path, etc.)
- When the user is given the choice to enter a building a different button should be used (something else instead of E)

U8: It should show where I am on the map anytime. So I can better navigate

The second part should allow me to go back to the text. So I can read it without the voice. That way the information will become more accessible (in case I forget any information)

U9:
a) Improve panoramic applications
b) Less information should be provided
c) Highlight in dialogue window should be correct (rollover is wanted but should work correctly, otherwise remove it completely)

Question 3: **Do you have any comments about the virtual guide?**

U1: See above

U2: No comments she is very realistic

U3:
- More realistic
- More natural voice less speed

U4:
- Very friendly
- Very sexy

U5: No

U6:
- See above
- More human in her behavior and not realism.
- More appropriate dressing

- She remind me more of a girl going out

U7: See above + I liked the gender I wouldn't think of anything else.

U8: It wouldn't make any difference if the character was more realistic

The guide was useful in showing objects in the environment. Zoom functionality and an arrow would still be the same.

U9: Should take less space on the screen

Question 4: **Do you think there is any way to improve your experience with the virtual guide?**

U1: See above

U2: I think the design of the system doesn't leave the user to have any questions. The system is complete and its contents understandable.

U3: No

U4: No the guide is OK

U5: The second system is better because there was no guide and the subtitles

U6: See Above

U7:
- If the guide would have been just a face (without any gestures, etc.)
- I would prefer a simple pointer to show me objects in the environment!

U8:
- Realism wouldn't make a difference
- I am mostly interested for the information provided and not the guide. I prefer the text.
- A more natural mode of communication (speech recognition) wouldn't make a difference

U9:
a) Speech recognition is necessary
b) A small rewind button (to rewind the character a few seconds only) – only in the descriptions
c) Time Counter to show the user the duration of the presentation. So the user will know how long the system will speak.

Question 5: **Do you have any comments on the questionnaires you filled-in? (For example would you like to expand on any underlying issues found in the questionnaires?)**

U1: See above

U2: No I think it's a good questionnaire

U3: No

U4: No

U5: No

U6: Questionnaires should be more balanced and with numbers so the user won't have to count the boxes in order to decide

U7: No

U8: No

U9: No comments

Interviews - Order 2 (Absent vs. Present):

Question 1: **Do you have any comments about the systems?**

U10:
- The first system was much easier. The second system was much nicer.
- In the second system it was very difficult to synchronize what the system said with the panoramic application.
- It was very difficult in the second system to hold any information. The guide attracted my attention.

U11:
- I like the general idea
- The second system is more confusing. The guide attracted my attention from the presented information. I would mostly prefer an arrow to show me around
- I mostly like the first system

U12
- The second part was not of interest to me as the content is not relevant to my interests
- In the second system the guide spoke too fast
- It was very difficult to synchronize between the system and the panoramic applications
- Panoramic applications were not that clear

U13: I did not like the guide, not her appearance but rather her pronunciation was wrong. The pronunciation sometimes made it difficult for me to understand some words. Her speech was not natural, especially in words that were multi-syllabus.

U14:
- The user should be able to choose all information scenarios and not just one
- The user should not follow the guide but the guide should help the user. At the beginning, the guide should give an explanation why she is there. I could have two ways of touring
- Free tour. Where the user would have the choice to walk around freely. In this choice you don't have pre-constructed text but rather buttons in the panoramic indicating that there is information for the particular point. For example for the canon in the main square you could have a menu

indicating that there is Architectural, Historical, or other information for the particular point. Once the user selects what he wants the guide must sync and present the particular information. In the real environment the character would have a panoramic in its background to allow the user to scroll and discover the points for which the character knows about.

- Guided tour where the user will have to follow the guide (as it is now)
- In the panoramic I must have a green arrow showing which way. The pointing gestures of the avatar are not enough for correct guidance.
- Interface design. The construction of the dialogue menu is not sufficient. For example there are no bullet points to distinguish the questions. The dialogue menu appears more like a continuous text and not as selection of questions. A question mark button should be used where the user would click to make the questions appear
- An index section must be used to indicate which areas of the castle I can visit

U15:

- The avatar has to be more human. Her movements are not that realistic. Her movements are not synchronised with the audio. I like that when she explains she points
- Not good facial expressions (a bit irritating)
- The character gives me that human feeling. It's not just a computer giving me information.
- I was completely disoriented with the first system (text only). I didn't even see the canon at the centre of the square.

U16:

- In the second system I could not use the pause button and hence I don't think that my retention performance was like the first system.
- Speech in the second system was faster than in the first. As there is no text for me to read, you should lower the rate of the speech considerably.
- Generally I am pleased with the whole experience. I paid more attention to both systems (the avatar was more interactive) compared to a book. I remember more things. I think I paid more attention to the avatar, and hence I did better in the retention test (I am a visual person)

U17:

- The content has the proper length. The extra words for the navigation (e.g., been polite) were not necessary. The comments that have to do with using the system could become sorter.
- In the dialogue window, you could add a sort audio to hear the questions. This will make the dialogue more natural.
- I would like the objects at each location (e.g., the marbles of the churches) to become clickable (only the most important objects). This would allow me to keep more information in mind. The content could either a) become more brief and to be transferred in the clickable objects b) to be repeated from the objects.
- As a presentation method, none of the two systems is better

U18:

- The speech for both systems should be less continuous and with more pauses.

- It was not clear what I could do with each of the buttons

Question 2: How do you think I can improve the design of the system?

U10:
- The first system was OK
- Slower provision of the information

U11:
- Improved panoramic applications
- Less information is desirable (a summary of the existing ones)

U12:
- The guide attracted my attention too much, she has a very intense presence
- The buttons in the panoramic applications should not be that close

U13:
- Natural voice (it becomes more natural to the user)

U14:
- Dynamic adaptation of the panoramic images in the background of the character based on the location of the user.
- To be able to learn more information about arbitrary places in the castle. Each castle must do its own panoramic so the user could choose the points for which he wants to hear about (see above menus in the panoramic).

U15:
- You have to either move with the text or the buttons. One of the methods should be there.
- Panoramic should have a drag pointer (change the pointer when you drag)
- Panoramic should have different speeds when you are in the middle of the panoramic and different when you are at the edge.

U16:
- Speak in lower speed
- Better aesthetics in the absent system
- Better dialogue windows (e.g., enable multi-touch to be able to scroll with your finger etc.)

U17:
- Better hardware (with the new devices, multi-touch, etc.)
- See above

U18:
- I would like to see more movement in the background (e.g, to zoom to the points which the system is providing information about)
- The windows should not be permantely visible on the interface but should become available only when the user needs them.
- The design of the system should be more modern (e.g., with more vibrant colours)

Question 3: **Do you have any comments about the virtual guide?**

U10:
- Appearance. Make her more Greek.
- Increased realism (like the avatar movie) would make a difference – just for the effect (mouth)

Non-realistic lip-sync

U11:
- Even a guide with increased realism wouldn't make any difference

U12:
- The guide should not be any more realistic. You should include more attractive information in both systems (e.g., for shops, reconstructed houses, etc.)

U13:
- Her lip-sync was not natural
- I didn't notice her body language at all.

U14:
- The virtual guide is not necessary for the system. I could only have the voice and an arrow to highlight the points for which she is speaking about

U15:
- See above. If the guide was like the avatar movie (in terms of realism and behavior) I would have remembered more information. The WOW factor (for the graphics) would have made me to pay more attention and hence learn more. Even with the existing avatar if her behavior was more human it would have made a huge difference in my retention of the presented information.

U16:
- I liked the virtual guide and her movements

U17:
- I am not someone who will look at the graphics
- If the guide was at the quality of the avatar movie it could have been better. However, the avatar would have attracted my attention more.

U18:
- The guide was overall ok but her movements were too intense

Question 4: **Do you think there is any way to improve your experience with the virtual guide?**

U10:
- Increased realism
- More realistic lip-sync

Less speed in the body language

U11:
- Better lip-sync
- Remove the guide completely

U12:
- I cannot imagine a way.

U13:
- Even if the avatar was at the range of realism of avatar movie it wouldn't make a difference to my retention of the presented information

	▪ However the avatar was not a problem to me. Her presence didn't affect me. An arrow showing the points of interest would still make the same job.
U14:	▪ Even if I could have an avatar in the realism of the movie it wouldn't make any difference to the retention of information. It attracted my attention too much. Speech recognition and natural language processing in the system that doesn't have the avatar. Even if her behavior was human it wouldn't make any difference
U15:	▪ Better body language, and then better graphics
U16:	▪ See above plus a more personalized greeting (e.g., hello Maria) ▪ When I take a wrong turn at the panoramic, it would be nice if the avatar would say "No Maria you took the wrong direction, please try again" ▪ If the avatar would have been like the movie "Avatar" the interaction would have been more effective, as the avatar would have been more credible. However, most likely it would have not been more effective in giving information.
U17:	▪ No, I want the user to be involved in the interaction. ▪ The avatar is good because you see an anthropomorphic character and it is familiar. But you know it is not a real person so it is not relevant.
U18:	▪ The guide should not be that expressive and should be less in the camera. ▪ The guide should not stare the user constantly in the eyes. ▪ The guide should have a smaller size.

Question 5:	**Do you have any comments on the questionnaires you filled-in? (For example would you like to expand on any underlying issues found in the questionnaires?)**	
U10:	No	
U11:	Very long and detailed questions	
U12:	No	
U13:	Spelling mistakes Some of the questions are repeated	
U14:	No	
U15:	The system is still at the beginning. The questionnaires are not very effective.	
U16:	Well Structured	
U17:	Easy to answer	
U18:	None	

Table D.1.10: Post-task interviews in experiment one

Retention Test

Please fill-in the visual questions first, and then the questions of simple text. At each question please rate the confidence of your answer in a scale of 1-10 (1 = completely at random, 5 = not so confident, 10 = totally confident)

Visual Questions:

What is that and for what it was used for? (Confidence =)

What is that and for what it was used for? (Confidence =)

What is that and for what it was used for? (Confidence =) (Architecture Scenario Only)

What is that and for what it was used for? (Confidence =) (All scenarios)

What is that and for what it was used for? (Confidence =) (All scenarios except architecture)

What is that and for what it was used for? (Confidence =)

What is that and for what is its origin? (Confidence =)

Textual Questions (12 Questions/Scenario)

1) The word Monemvasia means _____ (Confidence =)

2) The face of the main gate is made of _____ (Confidence =)

3) The hemisphere that sits just above the gate opening is itself flanked on each side by _____ (Confidence =)

Architecture Scenario ONLY (One question)

4) Above the facing of the gate, to the left of the corbelling, are the remains of _____, identifiable by the small pieces _____(Confidence =)

History Scenario ONLY (One question)

4) The larger section of the still visible defense system of the lower town was built at _____ (Confidence =)

Biographical Scenario ONLY (One question)

4) The main gate of the castle was built during the _____ (Confidence =)

Architecture Scenario ONLY (One question)

5) The first storey of Yannis Ritsos house was used as _____(Confidence =)

6) The second storey of Yannis Ritsos house has a _____, and _____ with _____ (Confidence =)

7) The courtyard of the house of Yannis Ritsos was probably covered by _____ (Confidence =)

8) The roof of Yannis Ritsos house is consisted of large curved tiles arranged in rows side by side. The first row is with the _____ and the second row is with the _____ (Confidence=)

All the other Scenarios (Three Questions)

5) Yannis ritsos spent the first years of his life at _____(Confidence =)

6) The inscription at the base of the Yannis Ritsos bust says _____ (Confidence =)

7) Yannis Ritsos died on _____ in _____(Confidence =)

Biographical Scenario ONLY (One question)

8) Yannis Ritsos on 1936 wrote the _____ and on 1966 the _____ two milestone collections of his poetic contribution (Confidence =)

History Scenario ONLY (One question)

8) Yannis Ritsos returned to Monemvasia on _____ after 20 years where he wrote the poems _____ and _____ (Confidence =)

---

9) The local name of the main square of the castle is _____ (Confidence =)

10) The square building on the main square of the castle is used as a _____ (Confidence =)

11) Straight from the main square of the castle, the view opens up to the lower-lying parts of the lower town, and far off to the southwest _____ disappear on the horizon (Confidence =)

Architecture Scenario ONLY (One question)

12) The two other big squares of the lower town, called _____ are located in the lower area of the town (Confidence =)

Biographical Scenario ONLY (One question)

12) Above the entrance of the Episcopal residence is built a _____ that depicts _____ (Confidence =)

History Scenario ONLY (One question)

12) The museum was built during the _____ (Confidence =)

Table D.1.11: The retention test used in experiment one

Experiment Two:

Participants:

Order	Participant	Age/Gender	Profession	Mobile device user
S/T	001	35/Female	Opera Singer	2G phone /laptop
S/T	002	37/Male	Sales Manager	3G phone/laptop
S/T	003	32/Male	Web Designer	2G phone
S/T	004	32/Male	Student	2G phone
S/T	005	38/Female	Administrator	3G phone
S/T	006	34/Male	Administrator	3G phone
S/T	007	30/Female	Waitress	3G phone
T/S	008	32/Male	English Teacher	3G phone
T/S	009	32/Female	Marketing Manager	3G phone
T/S	010	32/Female	Student	2G phone
T/S	011	37/Male	Company Director	2G phone
T/S	012	38/Male	Electrician	3G phone
T/S	013	33/Female	Photographer	2G phone
T/S	014	33/Male	Insurance Agent	3G phone

Table D.2.1: Participants in experiment two

Objective Assessment:

Present (%) (N=14)	**Absent (%) (N=14)**	**Order**
12	38	S/T
12	14	S/T
36	29	S/T
24	19	S/T
40	19	S/T
40	29	S/T
12	5	S/T
9	62	T/S
10	29	T/S
11	71	T/S
12	29	T/S
13	29	T/S
14	5	T/S
15	33	T/S
18.5	**29.3**	**Mean**

Table D.2.2: Participants' retention scores in experiment two

Subjective Assessment:

Present (N=14)	**Absent (N=14)**	**Order**
4	2	S/T
5	3	S/T
3	4	S/T
4	3	S/T

1	1	S/T
5	5	S/T
4	3	S/T
5	4	T/S
3	2	T/S
4	2	T/S
4	5	T/S
3	4	T/S
5	3	T/S
4	2	T/S
3.8	**3.0**	**Mean**

Table D.2.3: Participants' difficulty ratings in experiment two

Questions	**S/T (n=14)**		**T/S (n=14)**		**AVG**
	P	**A**	**A**	**P**	**P/A**
1) The information task is too complex	2	1	3	4	3/3
2) The information task is difficult to learn	2	1	3	3	3/2
3) The process of extracting information about a location from the system is too difficult to learn	2	2	4	3	3/3
4) The screens are inconsistent with the provided information about a location	2	2	3	2	3/2
5) The completion of the information task requires high precision (e.g., to photograph a QR-Code about a location)	3	3	4	3	4/3
6) The completion of the information task requires too much self-organization	3	3	5	5	4/4
7) I cannot really see what is on the screen of the system	1	1	1	1	1/1
8) I cannot really hear the speech of the system	1	1	1	2	1/2

9) The modalities used by the system (i.e., speech, images, gestures and face expressions) are too many and confuse me. I would prefer a simpler system (e.g., one with speech only and/or text)	1	2	2	3	2/3
10) It is difficult to make sense of the speech output of the system	2	2	3	3	3/3
11) The information provided by the system is poorly presented (too brief or too long)	2	2	3	3	3/2
12) The output the system (i.e., audio, gestures, face expressions and images) is poorly timed	1	2	2	2	2/2
13) The output of the system (i.e., audio, gestures, face expressions and images), is unclear	1	2	2	1	2/2
14) Some outputs of the system do not match the actual environment of the castle.	2	2	1	1	2/2
15) I need more detailed help from the system on how to photograph the QR-Codes	1	1	1	2	1/4
16) I find it hard to remember information about a location after it was presented by the system	6	4	5	4	6/3
17) I have to hold too much information in mind when using the system	5	3	5	3	5/2
18) I find it difficult to remember that I have to photograph a QR-Code to listen to a presentation. I would prefer a more automatic method	1	1	1	2	1/4
19) The system should prompt me to pay more attention to a presentation about a location	4	4	5	4	5/5
20) The system should respond appropriately (e.g., by pausing a presentation) when I am confused or overloaded with information	5	5	4	4	5/2
21) The system is too frustrating to use	1	2	1	1	1/2
22) The system is too annoying to use	1	2	1	1	1/3
23) The design of the system is not serious enough	1	3	1	2	1/5
24) The system is fun to use	6	5	6	5	6/2

25) The design of the system makes it difficult to learn what I need to learn to use it properly	2	1	2	2	2/3
26) It's hard to learn any of the information presented by the system	5	4	3	2	4/3
27) The information provided by the system should relate better to what I already know	3	4	4	2	4/4
28) The information provided by the system should relate better to my personal interest	4	4	4	4	4/3
29) I find it difficult to follow the information presentation about the four location (e.g., because of absence of structure, complicated terms or other reasons)	3	3	4	3	4/3
30) The structure of the information in each of the locations was not presented appropriately	2	3	2	2	2/3
31) The modalities used by the system (i.e., audio, images, gestures and face expressions) prevent me from constructing a clear "mental picture" of the information presented about each location.	2	3	2	3	2/4
32) A simpler style of presentation would have enable me to remember more about the castle	3	3	2	4	3/2
33) The allowances for my response errors (e.g., when I don't photograph a QR-code correctly) are not satisfactory	3	1	1	2	2/3
34) The system does not provide me with sufficient information to respond with appropriate reactions to its requests (what to do next)	2	3	2	2	2/2
35) The system requires me to make unreasonable responses (e.g., navigate hard-to-walk routes)	2	1	2	2	2/3
36) I make a lot of response errors with the system (i.e., wrongly retained information)	4	4	4	2	4/4
37) The electronic map is extremely simple to help me find the locations in the castle.	4	5	4	3	4/4
38) The system requires me to find locations in the castle that are very difficult to find.	3	3	2	2	3/3

39) I always have to seek the experimenter's help to proceed from landmark to landmark.	1	1	1	1	1/1
40) The system gives me no support to learn the information presented about a location (e.g. to repeat part of the presentation)	6	4	5	4	6/4

Table D.2.4: Mean responses to the workload questionnaire in experiment two

Retention Test (Simple Content)

If you find a question not clear enough, please ask for my help. At each question, please rate the confidence of your answer on a scale of 1-10 (1 = completely at random, 5 = not so confident, 10 = totally confident).

1) At the citadel of the castle you will find the _____ which connected it with the _____ (confidence =)

2) To reach the upper town of the castle you will have to walk _____ and go through the _____ (confidence =)

3) The castle has access to the sea through the _____ (confidence =)

4) The original building phase of the church of "Christ Elkomenos" most probably dates to _____ and its present architectural form is the result of _____ (confidence =)

5) An important heirloom from the church of "Christ Elkomenos" now exhibited in the Byzantine and Christian Museum in Athens is the _____ (confidence =)

6) Some of the post-byzantine portable icons that adorn the church of "Christ Elkomenos" today are _____ (write as many as you can remember) (confidence =)

7) The church of Panagia Myrtidiotissa is also known as _____ because it _____ (confidence =)

8) The worship of the "Panagia Myrtidiotissa" is associated with _____ (confidence =)

9) According to local lore, the church of "Panagia Myrtidiotissa" was founded by the _____ (confidence =)

10) Local tradition has it that the church of "Hagios Nikolaos" never _____ (confidence =)

11) The coat-of-arms above the entrance of the door of the church of "Hagios Nikolaos" shows a _____and belongs to _____ (confidence =)

12) The church of church of "Hagios Nikolaos" also functioned as _____, where the poet _____ was a pupil. (confidence =)

Table D.2.5: The retention test used in experiment two

***Open Comments (Both Orders)*:**

U1:
- Some sentences were too long
- There should be a choice for repeating certain parts (something like reverse)
- There should be a choice to turn of the voice
- There should be a window to make the whole text of the description available to the user
- There should be a way to highlight on the specific parts of the attraction to which the content refers to
- In some of her movements the avatar was distracting and disorienting
- When photographing the QR-Code sometimes I could hear the click of the camera and sometimes not

U2:
- Way too much information – details
- I would like a better questionnaire. I think some of the questions were not clear enough
- Who is explaining the way to use the system? Will there be a manual?

U3
- The information provided by the system about the architecture of the buildings is complex and cannot be memorized easily. I would prefer the system to guide me how to go the next attraction. Otherwise the system

was correctly structured and with a reasonable continuation.

U4:

- The information provided by the present system is difficult.
- The photographs cannot be easily combined with the provided information
- The absent system was clearer from the first though the avatar was missing. Her presence improves the aesthetics of the interface but not the practical part of the application
- I would like to use the pause button anytime
- I had a great time testing the system. Thank you very much

U5:

- The system that was made by Yannis was very good with an excellent and very easy to use design. The information provided were excellent and very enlightening, as well as the pictures that were displayed during the narration. I had some problems photographing the QR-Code but overall I felt that the system is a new way of information presentation and I think it will succeed in its applications.

U6:

- Way too many dates
- Way too many churches
- Way to many architectural details to remember

U7:

- The system includes information that is useless to a potential visitor of the castle. That's because it refers to elements that would be useful to someone more specialized, like for example, architecture, structure of buildings, etc.
- Information that contains too many technical details is tiresome for someone that visits a tourist attraction.
- Then during the description of the church, it would be nice to highlight the points for which the system is providing information for.
- Consider providing alternative images (e.g., something like a virtual tour)
- I found problems in the construction of the sentences that gave the wrong impression. Very big sentences are tiresome and decrease my interest to the information that is given each time.

- You need to be careful in the construction of the questionnaires. The questions need to be clear and simple.

U8:

- A well designed system, simple enough to use. I would like less information for the construction of the churches and more information about the history of each monument-point. It would be nice not to speak everything at once. Consider allowing the user to use a pause button or decrease the rate of the avatar's speech

U9:

- No comments

U10: The system is fun in its presentation and very simple in its use. I prefer the method of presentation with the avatar. I think the method with the text reading and the synchronous presentation of pictures is tiresome. The information that was given to me was difficult to comprehend and memorize. The rate of the avatar's speech was too fast. I would prefer a slower speech for easier memorization. Finally I would suggest the use of more pictures at each personation and the use of videos.

U11: The guide speaks with gaps. In addition, it gives too concentrated information without many gaps and pauses. The picture was quite good and the system was easy to use. Finally the avatar is very attractive.

U12:

- The girl's hair moved too often
- Spastic movements
- Acoustics not very clear at some points
- Said period at the punctuation point
- I suggest have system to ask what you want to get out of Monemvasia (e.g., historic facts, architecture, cultural, interesting facts, etc.)
- Zoom in option to see clearly the artifacts
- Question 17 (on the 1-7 Questionnaire) not clear whether you are talking about the software instructions or the historic information given

U13: Both systems were quite entertaining and easy to use. The only disadvantage is that it is required by the user to be increasingly focused for memorizing the provided information.

U14: The system is interesting and practical to use. In order to make it easier for the user to follow its contents, it could zoom on to the different artifacts for which it is providing information about.

Table D.2.6: Open comments in experiment two

Experiment Three:

Order	Name	Age/Gender	Profession	Mobile device user
S/C	001	20/Male	Student	3G phone
S/C	002	23/Male	Student	3G phone/laptop
S/C	003	20/Male	Student	3G phone/laptop
S/C	004	23/Male	Student	3G phone/laptop
S/C	005	21/Male	Student	3G phone/laptop
S/C	006	18/Female	Student	3G phone/laptop
S/C	007	20/Male	Student	3G phone/laptop
S/C	008	21/Female	Student	3G phone/laptop
S/C	009	28/Male	Student	3G phone/laptop
C/S	010	29/Male	Student	3G phone/laptop
C/S	011	19/Male	Student	3G phone/laptop
C/S	012	20/Male	Student	3G phone/laptop
C/S	013	21/Male	Student	3G phone/laptop
C/S	014	22/Male	Student	3G phone/laptop
C/S	015	22/Female	Student	3G phone/laptop
C/S	016	23/Male	Student	3G phone/laptop
C/S	017	23/Male	Student	3G phone/laptop
C/S	018	27/Male	Technician	3G phone/laptop

Table D.3.1: Participants in experiment three

Objective Assessment:

Present(N=18)	Absent(N=18)	Order
888	1349	S/C
987	1230	S/C
956	1342	S/C
948	1253	S/C
1033	1088	S/C
932	1174	S/C
946	1088	S/C
889	1269	S/C
2210	1165	S/C
1224	848	C/S
1060	827	C/S
1287	796	C/S
1285	797	C/S
1114	746	C/S
1509	806	C/S
1098	854	C/S
1347	945	C/S
1093	773	C/S
1155.8	**1019.4**	**Mean**

Table D.3.2: Time taken (in seconds) to complete the tour in experiment three

Present (N=18)	Absent(N=18)	Order
2	4	S/C
2	1	S/C
3	2	S/C
3	4	S/C
3	1	S/C
2	4	S/C
1	4	S/C
6	5	S/C
1	6	S/C
1	4	C/S
3	4	C/S
5	2	C/S
1	0	C/S
2	4	C/S
3	2	C/S
2	2	C/S
5	4	C/S
0	1	C/S
2.5	**3**	**Mean**

Table D.3.3: Frequency of getting lost in experiment three

Subjective Assessment:

Present (N=18)	Absent (N=18)	Order
4	3	S/C
3	4	S/C
0	0	S/C
4	2	S/C
0	0	S/C
0	0	S/C
4	4	S/C
3	3	S/C
4	2	S/C
4	3	C/S
4	3	C/S
4	2	C/S
2	4	C/S
5	3	C/S
4	3	C/S
4	2	C/S
3	3	C/S
5	4	C/S
3.1	**2.5**	**Mean**

Table D.3.4: Participants' usefulness ratings in experiment three

Order	Present (Y/N) (n = 18)	Absent (Y/N) (n = 18)
S/C	6/0	6/0
S/C	5/1	6/0
S/C	6/0	6/0
S/C	6/0	6/0
S/C	5/1	4/2
S/C	5/1	6/0
S/C	5/1	3/3
S/C	5/1	5/1
S/C	6/0	6/0
C/S	4/2	4/2
C/S	4/2	6/0
C/S	4/2	6/0
C/S	5/1	6/0
C/S	5/1	5/1
C/S	6/0	5/1
C/S	4/2	2/4
C/S	5/1	4/2
C/S	6/0	6/0
Total:	**92/16**	**92/16**

Table D.3.5: Total responses to the object recognition (Yes/No) questionnaire in experiment three

Questions	S/C (n=18)		C/S (n=18)		AVG
	P	A	A	P	P/A
1) The navigation task is too complex	2	3	4	3	3/3
2) The navigation task is difficult to learn	2	2	3	2	3/3
3) The process of extracting navigation instructions from the system is too difficult to learn	3	3	3	3	4/3
4) The screens are inconsistent with the provided navigation instructions	4	3	3	3	4/4
5) The completion of the navigation task requires high attention skills (e.g., for not losing a landmark)	5	4	4	5	3/4
6) The completion of the navigation task requires too much self-organization	3	4	4	3	2/3
7) I cannot really see what is on the screen of the system	2	2	3	2	2/2
8) I cannot really hear the instructions of the system	2	3	3	3	3/2
9) The modalities used by the system (i.e., speech, images, gestures and face expressions) are too many and confuse me. I would prefer a simpler system (e.g., one with speech only and/or text)	3	3	4	2	3/3
10) It is difficult to make sense of the instructions used by the system	2	3	4	2	2/3
11) The navigation instructions provided by the system is poorly presented (too brief or too long)	2	2	4	3	2/2
12) The output the system (i.e., audio, gestures, face expressions and images) is poorly timed	2	2	3	2	3/2
13) The output of the system (i.e., audio, gestures, face expressions and images), is unclear	3	3	3	3	2/3
14) Some outputs of the system do not match the actual environment of the castle.	3	2	3	3	3/2

15) I need more detailed help from the system on how to properly locate landmarks	3	3	4	3	4/3
16) I find it hard to remember a navigation instruction after it was presented by the system	3	4	4	3	3/4
17) I have to hold too much information in mind when using the system	2	3	4	3	2/3
18) I find it difficult to remember that I have to tap a button to get the next instruction. I would prefer a more automatic method	2	2	2	2	4/2
19) The system should notify me that a landmark is approaching before I get there	5	5	4	5	6/4
20) The system should remind me when I get off track.	6	6	5	5	2/6
21) The system is too frustrating to use	2	2	3	2	2/2
22) The system is too annoying to use	2	2	3	2	1/2
23) The design of the system is not serious enough	2	2	2	2	5/1
24) The system is fun to use	5	5	4	5	1/5
25) The design of the system makes it difficult to learn what I need to learn to use it properly	2	2	3	2	2/1
26) It's hard to learn the route presented by the system	3	3	4	2	3/2
27) The routes provided by the system should relate better to my personal interests	3	3	2	2	5/3
28) The system should allow routes tailored to the castle's environmental context (e.g., the easiest to follow route, the shortest route, etc)	5	5	5	4	4/5
29) I find it difficult to construct a mental map of the route in the castle as it is presented by the system.	3	4	3	3	3/4
30) The structure of the route is not presented well	2	2	4	2	4/3

31) The modalities used by the system (i.e., audio, images, gestures and face expressions) prevent me from building a clear "mental map" of the route in the castle.	3	3	4	2	6/4
32) A simpler route would have been more enjoyable and easier to actually learn	3	5	3	3	5/6
33) The system should make allowances for my response errors (if I get too far of course)	5	5	5	5	3/5
34) The system does not provide me with sufficient information to respond with appropriate reactions to its requests (where to go next)	3	3	3	3	3/3
35) The system requires me to make unreasonable responses (e.g., navigate hard-to-walk routes)	2	3	3	3	3/3
36) I make a lot of response errors with the system (i.e., wrong navigation decisions)	3	3	3	2	3/3
37) The system requires me to find landmarks that are too difficult to find	2	3	3	2	2/3
38) I never know the correct navigational instructions in order to get to my destination	2	2	3	1	2/2
39) I always have to seek the experimenter's help to proceed from landmark to landmark.	3	2	3	2	4/2
40) The system gives me no support to learn the navigation instructions provided about the route (e.g. to repeat the instruction before a turn)	4	4	3	2	3/4

Table D.3.6: Mean responses to the workload questionnaire in experiment three

Comments - Order 1 (Simple vs. Complex):

U1: The system with the avatar was better, as it would show you where to go.

U2: No Comments

U3 No comments

U4: System A:

It is so much better when you are given a set of direction and there is someone who can describe with their body movements where to go to.

System B:

Reading a text means you need to be fast especially when it is a long one and sometimes you might not understand what step you really need to take. I prefer system A compared to system B

U5: No comments

U6: No comments

U7: There is no difference between the present and absent systems.

U8: The avatar spoke a lot of the instructions at once so it was hard to remember what she said. Furthermore, the videos are not very clear, so it's hard to relate the system with the Videos. She preferred the system with the subtitles, because she could read in case she wouldn't understand.

U9: The absent version talked too fast. Some words are not clear and the absent delivered most instructions at the same time. The avatar was better because it showed you where to go. However, it would be best for the avatar if she could use a real voice instead of an artificial one.

Comments - Order 2 (Complex vs. Simple):

U10: The avatar-based system directed you which way to go

U11: The virtual guide made it more interactive. You cannot use return in the second system, but it is less hard.

U12: System A is easier because of the nature of instruction

U13: Since the avatar was not there, it was easier to concentrate on the given instructions

U14: Difficult to read and listen (subtitles made it difficult to understand instructions and they were too fast)

U15: More difficult to read and see on the screen

U16: It was clearer on the first one. The audio was not that distorted. Text attracted attention away. The avatar made it more user friendly but had no effect on the instructions provided.

U17:

- The avatar on the present system was useful in giving directions but it should include text not just voice and gestures for giving directions.
- In the absent system text was good but it was annoying as I was to read while watching the videos to decide where to go.

	• In addition, the voice in the absent system was not clear. • Some of the videos on the laptop were not clear
U18:	• System A: The avatar is better as it describes the routes with relevant gestures • The background has to be up to date with the video • System B: avatar was better at pointing but having the written text you can look back if you forgot something

Table D.3.7: Comments in experiment three

APPENDIX E:

This appendix contains several tables. The tables contain data from the empirical studies that were conducted to evaluate the last three of the six prototype mobile tour guide systems. Full details of the analysis conducted on the data and suggestions can be found in Chapter 8.

Experiment Four:

Participants:

Order of task	Tester Code	Age/Gender	Profession	Mobile device user
First route vs. Second Route	001	22/Male	Student	3G phone/laptop
First route vs. Second Route	002	23/Male	Student	3G phone/laptop
First route vs. Second Route	003	22/Male	Student	3G phone/laptop
First route vs. Second Route	004	53/Male	Student	3G phone
First route vs. Second Route	005	21/Male	Student	3G phone/laptop
First route vs. Second Route	006	26/Male	Student	3G phone/laptop
Second Route vs. First Route	007	24/Male	Student	3G phone
Second Route vs. First Route	008	21/Male	Student	3G phone
Second Route vs. First Route	009	25/Male	Student	3G phone/laptop
Second Route vs. First Route	010	23/Male	Student	3G phone/laptop

Second Route vs. First Route	011	25/Male	Student	3G phone/laptop
Second Route vs. First Route	012	28/Male	Student	3G phone/laptop

Table E.4.1: Participants in experiment four

Objective Assessment:

System A (%) (n=12)	System B (%) (n=12)	Order
12	0	First route vs. Second Route
3	4	First route vs. Second Route
6	10	First route vs. Second Route
10	8	First route vs. Second Route
16	12	First route vs. Second Route
48	31	First route vs. Second Route
6	6	Second Route vs. First Route
4	1	Second Route vs. First Route
14	10	Second Route vs. First Route
0	0	Second Route vs. First Route
6	3	Second Route vs. First Route

10	7	Second Route vs. First Route
7	4	Second Route vs. First Route
6	2	Second Route vs. First Route
11	**7.8**	**Mean**

Table E.4.2: Participants' retention scores in experiment four

Questions (Q=60)	Scripts	Shallow parsing	Deep Syntactic Processing
1	0	0	20
2	5	5	5
3	20	20	5
4	0	0	5
5	20	20	5
6	10	10	5
7	20	20	10
8	5	5	5
9	20	20	20
10	20	5	20
11	5	5	20
12	20	20	20
13	10	10	5
14	5	5	5
15	20	20	0
16	5	5	0
17	20	20	0

18	0	0	0
19	5	5	5
20	20	20	5
21	20	20	20
22	0	0	5
23	20	20	20
24	20	20	0
25	5	5	5
26	20	20	20
27	5	5	0
28	5	5	5
29	5	5	5
30	20	20	0
Total (A-C)	**350/58%**	**335/56%**	**240/40%**

Table E.4.3: Algorithmic Comparisons per locations (Location A to Location C)

Questions	Scripts	Shallow parsing	Deep Syntactic Processing
31	5	5	0
32	5	5	0
33	20	20	20
34	20	20	5
35	20	20	5
36	0	0	0
37	5	0	5
38	5	5	5
39	20	20	5
40	10	10	0
41	5	5	0
42	10	10	10
43	20	20	20
44	5	5	5
45	20	20	20
46	5	5	5
47	0	0	0

48	10	10	5
49	5	5	0
50	20	20	5
51	20	20	5
52	20	20	20
53	5	0	0
54	10	10	0
55	10	10	0
56	20	20	5
57	20	20	0
58	10	10	5
59	20	20	0
60	10	10	0
Total (D-F)	**355/59%**	**345/57%**	**150/25%**

Table E.4.4: Algorithmic Comparisons per locations (Location D to Location F)

Subjective Assessment:

Measures	Scripts / Parsing (n=12)		Parsing / Scripts (n=12)		AVG
	Scripts	Parsing	Parsing	Scripts	S/P
Clarity	6.8	6.1	6.7	6.6	6.7/6.0
Wording	6.5	6.1	6.8	6.2	6.3/6.0
Sense	6.3	5.8	6.8	6.0	6.1/5.7
understandable	6.8	6.3	7.1	6.5	6.6/6.2
Simplicity	6.6	6.8	6.7	6.6	6.4/6.6
Fun	5.8	5.0	6.7	6.5	5.9/5.2
annoying	2.2	2.9	2.3	2.4	2.2/2.6
interesting	5.9	4.7	6.6	6.5	5.8/4.9
intelligent	6.1	5.1	6.6	7.0	6.0/5.3
Stimulating	5.2	4.5	5.9	5.8	5.2/4.7
tiresome	2.3	2.5	4.1	4.0	2.2/2.4
unpleasant	2.0	2.2	2.8	2.9	1.9/2.0
Accuracy	6.1	5.2	7.0	6.5	6.0/5.3

Table E.4.5: Mean responses to the answers-impression questionnaire

Experiment Five:

Order	Participant	Age/Gender	Profession	Mobile device user
S/C	001	23/Male	Student	3G phone/laptop
S/C	002	24/Male	Student	3G phone/laptop
S/C	003	25/Female	Student	3G phone
S/C	004	22/Male	Student	2G phone
S/C	005	24/Male	Student	3G phone/laptop
S/C	006	25/Male	Student	2G phone/3G phone/tablet
C/S	007	24/Male	Student	2G phone/3G phone/tablet
C/S	008	25/Male	Student	3G phone/laptop
C/S	009	26/Male	Student	3G phone/laptop
C/S	010	22/Male	Student	3G phone
C/S	011	26/Male	Student	-
C/S	012	31/Male	Student	3G phone/laptop

Table E.5.1: Participants in experiment five

Subjective Assessment:

Order	Full (Y/N) (n = 12)	Low (Y/N) (n = 12)
S/C	8/7	12/3
S/C	11/4	14/1
S/C	5/10	9/6
S/C	7/8	11/4
S/C	7/8	11/4
S/C	9/6	10/5

C/S	11/4	10/5
C/S	11/4	4/11
C/S	12/3	9/6
C/S	11/4	7/8
C/S	10/5	8/7
C/S	9/6	8/7
Total:	111/69	113/67

Table E.5.2: Object recognition questions in experiment five

Full (n=12)	**Low (n=12)**	**Order**
2	4	S/C
4	3	S/C
2	3	S/C
4	3	S/C
3	5	S/C
2	4	S/C
5	3	C/S
4	3	C/S
5	3	C/S
4	2	C/S
5	2	C/S
4	1	C/S
3.6	**3.0**	**MEAN**

Table E.5.3: Participants' usefulness ratings in experiment five

Full (%) (n=12)	Low (%) (n=12)	Order
11	7	S/C
11	21	S/C
0	14	S/C
17	21	S/C
11	7	S/C
6	21	S/C
28	14	C/S
28	50	C/S
22	14	C/S
33	14	C/S
33	29	C/S
17	0	C/S
18.0	**17.6**	**MEAN**

Table E.5.4: Participants' retention scores in experiment five

Order 1 (Full Competent vs. Low Competent):

U1:
- Repetition made the low-competent system easier to follow than the full competent guide. There was no different in terms of guide.

U2:
- The low-competent was more difficult to understand because of the voice. There was a difference between the two guides. The full competent was clearer.

U3
- The low competent system was better because it didn't use too many animations. Voice was better in the low competent too

U4:
- Compared to the full competent avatar, the low-competent avatar looks more real

U5:
- The content of the low-competent system is clearer

U6:
- With the full-competent system I am getting some acoustic problems. The fully-competent avatar was taking my focus off. The low-competent avatar is much more comfortable to use than the full-competent avatar because I can focus on the content.

Order 2 (Low Competent vs. Full Competent):

U7:
- The full-competent system was more interactive because of the guide gestures.

U8:
- There was more interaction from the full-competent guide (gestures, etc.)

U9:
- The full-competent avatar is clearer than the low-competent avatar. Gestures made me more visually aware what she was talking

U10:
- The low-competent avatar lacks of gestures to signify important information. The full-competent language was clearer. The way she was presenting was more interactive because of gestures.

U11:
- The guide wasn't really showing what she is talking about. Gestures made the full competent guide more visual.

U12:
- The full-competent guide is more descriptive (content + guide) than the low-competent guide

Table E.5.5: Comments in experiment five

Experiment Six:

Order	Tester Code	Age	Gender
S/C	001	20-30	Female
S/C	002	20-30	Female
S/C	003	20-30	Female
S/C	004	20-30	Male
S/C	005	20-30	Male
S/C	006	20-30	Male
C/S	007	20-30	Male
C/S	008	20-30	Male
C/S	009	38	Male
C/S	010	60+	Female
C/S	011	40	Female
C/S	012	35	Female

Table E.6.1: Participants in experiment six

Order	AG (Y/N) (n = 12)	NAG (Y/N) (n = 12)
S/C	8/7	12/3
S/C	9/6	12/3
S/C	8/7	14/1
S/C	9/6	10/5
S/C	13/2	9/6
S/C	13/2	5/10

C/S	7/8	9/6
C/S	8/7	10/5
C/S	7/8	9/6
C/S	4/11	7/8
C/S	5/10	9/6
C/S	9/6	11/4
Total	100/80	117/63

Table E.6.2: Object recognition questions in experiment six

AG (%) (n=12)	NAG (%) (n=12)	Type of Information
0	14	S/C
17	36	S/C
22	36	S/C
33	29	S/C
28	21	S/C
6	7	S/C
0	6	C/S
36	17	C/S
21	17	C/S
7	11	C/S
0	39	C/S
14	33	C/S
15.3	22.1	MEAN

Table E.6.3: Participants' retention scores in experiment six

AG(n=12)	NAG(n=12)	Order
4	2	S/C
1	1	S/C
4	3	S/C
5	3	S/C
3	4	S/C
4	4	S/C
4	4	C/S
3	2	C/S
3	4	C/S
4	4	C/S
3	2	C/S
4	2	C/S
3.4	**3**	**MEAN**

Table E.6.4: Participants' difficulty ratings in experiment six

Comments - Order 1 (3 Females (Simple) vs. 3 Males (Complex)):

U1:	Attention – grabbing	I did not know why she was yelling at me…I found it rude at the beginning but then it just made me laugh because she was right, I was not paying any attention. It was very difficult for me to understand the names of the places; it would be easier to remember them if you put their names in the picture.
	Non-attention grabbing	It is still difficult to remember all the details you ask in the questionnaire, although it was much easier for me to pay attention to the lady, so I could remember more things. Again it will be nice to have the names of the places written at the bottom of the image because all churches have similar and confusing names for me.
U2:	Attention – grabbing	In my opinion the recorded speech of the tour guide is extremely difficult to understand. In most cases I got distracted because I could not keep connected to the speech, and because names and dates she said were also difficult to remember. Also, the tour guides movements got me distracted from her presentation and got me thinking about other things or distracted.
	Non-attention grabbing	Even though the recorded speech seemed exactly as difficult as the previous one, the questions seemed a little bit easier because I was more concentrated than last time. But still I think it is quite difficult to understand. On the other hand, I prefer this system because the tour guide is a lot nicer, so I wanted to pay more attention to her than the one before.
U3:	Attention – grabbing	Not difficult, just a lot of history facts. Not so interested in that topic so couldn´t recall accurate information.
	Non-attention grabbing	Same as the A system…it was not difficult, though there were some terms that were hard to understand.
U4:	Attention – grabbing	I was distracted by the woman and its movements, another thing is that the topic is quite specific and the names are hard to remember. Some of the images that I had to later recognize were too small in the presentation I saw.
	Non-attention grabbing	After the first part of the test I was much more focused on exact names and dates and also more receptive to the images and all the small details on them.
U5:	Attention – grabbing	Is difficult to get information that probably you´ve never heard of from a machine (Is not a real person, sounds like a translator) it would be better and more personal to get somebody read the information. It will also get your attention as you feel it more personal. In the first presentation they don´t show the information they are talking about. Is better when they follow the icons or

		places with the avatar hands.
	Non-attention grabbing	This presentation is much easier because guides the user though the information they are talking about, is good when the avatar points the exact objects, icons and other that it´s explaining. On the other hand, is still difficult to fully understand something that you are not used to and a machine is explain, the voice speed is not natural and the level varies during the presentation. It would be much easier if a person speaks as the avatar.
U6:	Attention – grabbing	The videos have a lot of information to be remembered at once. I think that the speaking can be slower and the woman doesn't have to be there because it can be a distraction.
	Non-attention grabbing	The videos didn't have as much of information like in system A that was a good thing. I think that the topic was the one that didn't let me concentrate on the information given to me because I don't have a great knowledge of the words like the places or the people mentioned in the videos.

Comments - Order 2 (3 Males(Complex) vs. 3 Females(Simple)):

U7:	Attention – grabbing	During the test I was more interested in watching what the lady was doing and the images and places she was pointing, than in her speech. This made me forget a lot of information like names, shapes on the walls, inscriptions on the doors… etc. I think is an excellent program and hope this test can help with the investigation.
	Non-attention grabbing	In this exercise the lady didn't call me to pay attention to her, so it was easier to keep the interest on the objects and on what she was trying to explain to me. Like in the first experiment I think it's too much information to remember in a short time because she says names, dates and structural shapes of each church. I hope this test helps with the investigation.
U8:	Attention – grabbing	I didn´t like the way she was asking for my attention
	Non-attention grabbing	Too much information in too little time, sometimes boring
U9:	Attention – grabbing	The way the sound is reproduced makes the listening a little difficult at some parts as the voice of the lady sounds sometimes with some kind of interferences. Also, the Greek names, which are not familiar to any person who is not related with Greek culture or knows about it, makes you, sometimes, loose the focus of the tour and miss some of the given information
	Non-	Listening to the first part makes the person become more familiar with some

	attention grabbing	words and terminology, which enables the person to get some more information, although it cannot improve someone's memory, which is one of the most important reasons you can get wrong answers!
U10:	Attention – grabbing	I think the presenter takes over the presentation (she is too big on the screen) and she is a little bit distracting from the images of the churches, therefore, it makes it difficult for the person who is watching the presentation to remember all the details she is providing about the churches. Even though she is talking about the churches and providing information about them, the details of the churches are not easy to appreciate because the images are in the back of the screen
	Non-attention grabbing	I still believe that the presenter is way too big for the screen and she is distracting from the images of the churches and the details she is providing. I tried to shift my attention back to the images and pay attention to her speech but I found it difficult to stay focused without starting to think about the presenter.
U11:	Attention – grabbing	Too many presentations with many details. It is difficult to remember the details of each church. Some words are difficult to understand. The jokes did distract me from the speech
	Non-attention grabbing	Too many data to remember. It is difficult to remember the names, the centuries of the constructions and the information. There were too many presentations for a short period of time to retain the information.
U12:	Attention – grabbing	The avatar was very intrusive, and got most of the attention I was expected to pay to the churches. The voice tone was plain and sometimes it made me lose the attention. The jokes were a very good idea but didn´t pay off.
	Non-attention grabbing	It was a way easier with the shorter scenes. The Avatar is still intrusive and lost me at some points. I think more simple information is easier to remember (general).

Table E.6.5: Comments in experiment six

	ECA	Zone	Mean
Total_Time	Attention-Grabbing	background	54.8837
		Avatar	25.4947
		Total	40.1892
	Non-attention-grabbing	background	43.3308
		Avatar	19.2862
		Total	31.3085
	Total	background	49.1073
		Avatar	22.3904
		Total	35.7488
Number_Fixations	Attention-Grabbing	background	134.2500
		Avatar	70.6042
		Total	102.4271
	Non-attention-grabbing	background	89.5417
		Avatar	45.6458
		Total	67.5938
	Total	background	111.8958
		Avatar	58.1250
		Total	85.0104

Table E.6.6: Overall fixation data for experiment 6

3 Females – Order 1 (Simple Content)								
		Background		Avatar face				
User	Scene	Number of Fixations	Total Time in Zone	Number of Fixations	Total Time in Zone	Face Expression(s)	Interruption Face Expressions	Retention Test score
U1	Sc1	93	50.642	40	20.191	neutral/blank,relaxed	surprised, happy	0%
U1	Sc2	94	59.256	48	33.052	neutral/blank,relaxed,attentive	happy, skeptical, angry	0%
U1	Sc3	65	50.526	43	34.351	neutral/blank,relaxed,attentive	surprised, happy, angry	0%
U1	Sc4	131	46.182	87	28.156	neutral/blank,relaxed,attentive	happy, less happy	0%
U2	Sc1	94	46.594	48	20.47	neutral/blank,relaxed	slightly happy	17%
U2	Sc2	193	67.429	72	21.747	neutral/blank,relaxed	happy	17%
U2	Sc3	128	47.58	55	18.914	neutral/blank,relaxed	surprised, happy	17%
U2	Sc4	138	50.518	67	19.695	neutral/blank,relaxed	neutral/blank	17%
U3	Sc1	65	32.75	43	15.633	neutral/blank, happy,relaxed	neutral/blank	22%
U3	Sc2	145	52.179	89	27.395	neutral/blank,relaxed	happy	22%
U3	Sc3	110	37.712	81	24.512	neutral/blank,relaxed	happy	22%
U3	Sc4	113	41.945	93	29.224	neutral/blank,relaxed	happy	22%

Table E.6.7: Correlated data for the attention-grabbing ECA (females)

3 Females – Order 1 (Complex Content)							
		Background		Avatar face			
User	Scene	Number of Fixations	Total Time in Zone	Number of Fixations	Total Time in Zone	Face Expression(s)	Retention Test score
U1	Sc1	4	0	3	0	relaxed, happiness,neutral/blank,relaxed	29%
U1	Sc2	201	81.784	116	41.683	neutral/blank,relaxed	29%
U1	Sc3	113	51.682	83	31.356	neutral/blank,relaxed	29%
U1	Sc4	129	43.676	73	21.518	neutral/blank,relaxed	29%
U2	Sc1	110	57.862	37	18.447	neutral/blank,relaxed	24%
U2	Sc2	182	79.357	113	41.859	surprised,neutral/blank,sceptical[2]	24%
U2	Sc3	120	55.09	49	20.602	neutral/blank,relaxed	24%
U2	Sc4	115	42.113	68	21.308	neutral/blank,relaxed	24%
U3	Sc1	84	49.324	28	14.973	neutral/blank,relaxed	36%
U3	Sc2	148	68.971	79	31.101	neutral/blank,happy	36%
U3	Sc3	108	51.328	55	24.559	neutral/blank,relaxed	36%
U3	Sc4	60	32.283	34	14.666	neutral/blank,relaxed	36%

Table E.6.8: Correlated data for the non-attention-grabbing ECA (females)

3 Males – Order 1 (Simple Content)								
		Background		Avatar face				
User	Scene	Number of Fixations	Total Time in Zone	Number of Fixations	Total Time in Zone	Face Expression(s)	Interruption Face Expressions	Retention Test score
U1	Sc1	117	53.3	82	30.1	neutral/blank	Happy	33%
U1	Sc2	173	65.36	115	35.971	neutral/blank,relaxed	Happy	33%
U1	Sc3	153	54	97	29.4	neutral/blank,relaxed	slightly happy	33%
U1	Sc4	197	67.4	132	39.3	neutral/blank,relaxed	happy	33%
U2	Sc1	133	58.2	32	14.7	neutral/blank,frawn	neutral/blank	28%
U2	Sc2	202	68.369	31	8.717	neutral/blank,frawn	slightly happy	28%
U2	Sc3	147	51.1	69	20.3	neutral/blank,frawn	happy	28%
U2	Sc4	162	59.98	118	37.3	neutral/blank,frawn	neutral/blank	28%
U3	Sc1	107	50.3	83	31.68	neutral/blank,relaxed	neutral/blank	6%
U3	Sc2	185	68.9	108	36.7	neutral/blank,relaxed	neutral/blank	6%
U3	Sc3	131	52.7	78	26.82	neutral/blank,relaxed	slightly happy	6%
U3	Sc4	101	36.1	145	58.1	neutral/blank,relaxed	neutral/blank	6%

Table E.6.9: Correlated data for the attention-grabbing ECA (males)

3 Males – Order 1 (Complex Content)							
		Background		Avatar face			
User	Scene	Number of Fixations	Total Time in Zone	Number of Fixations	Total Time in Zone	Face Expression(s)	Retention Test score
U1	Sc1	71	46	12	12.977	neutral/blank,relaxed	29%
U1	Sc2	147	72.9	102	38.9	neutral/blank,relaxed	29%
U1	Sc3	91	49.9	38	21.18	neutral/blank,relaxed	29%
U1	Sc4	106	40.5	68	22.1	neutral/blank,relaxed	29%
U2	Sc1	102	54.5	18	9.348	neutral/blank,frawn	21%
U2	Sc2	153	70	49	22.1	neutral/blank,frawn	21%
U2	Sc3	95	43.5	55	20.6	neutral/blank,frawn	21%
U2	Sc4	125	42.2	78	23.6	neutral/blank,frawn	21%
U3	Sc1	126	60.2	17	11.179	neutral/blank, slightly happy,relaxed	7%
U3	Sc2	148	72.5	45	19.02	neutral/blank,relaxed	7%
U3	Sc3	108	49.3	55	20	neutral/blank,relaxed	7%
U3	Sc4	62	32.3	29	9.5	neutral/blank,relaxed	7%

Table E.6.10: Correlated data for the non-attention-grabbing ECA (males)

	3 Males - Attention_Grabing_Serious - Order 2 (Simple Content)							
		Background		Avatar face				
User	Scene	Number of Fixations	Total Time in Zone	Number of Fixations	Total Time in Zone	Face Expression(s)	Interruption Face Expressions	Retention Test score
U1	Sc1	170	72.464	73	26.156	neutral/blank,frawn	Curiosity	0%
U1	Sc2	268	87.56	91	28.2	neutral/blank,frawn	slightly happy	0%
U1	Sc3	197	62.498	111	33.192	neutral/blank,frawn	happy	0%
U1	Sc4	137	48.453	37	11.046	neutral/blank,frawn	happy	0%
U2	Sc1	62	24.56	35	11.766	neutral/blank,frawn	neutral/blank,frawn	36%
U2	Sc2	144	62.839	31	12.177	neutral/blank,frawn	happy	36%
U2	Sc3	101	46.405	22	7.266	neutral/blank,frawn	slightly happy	36%
U2	Sc4	111	38.126	5	2.512	neutral/blank,frawn,boredom	slightly happy	36%
U3	Sc1	110	59.429	58	23.515	neutral/blank,relaxed, attentive	neutral/blank	21%
U3	Sc2	197	69.704	136	42.045	neutral/blank,relaxed,attentive	slightly happy (smile)	21%
U3	Sc3	130	49.072	84	26.415	neutral/blank,relaxed, attentive	neutral/blank	21%
U3	Sc4	90	27.063	53	14.467	neutral/blank,relaxed, attentive	slightly happy (smile)	21%

Table E.6.11: Correlated data for the attention-grabbing ECA (males)

3 Males – Order 2 (Complex Content)							
		Background		Avatar face			
User	Scene	Number of Fixations	Total Time in Zone	Number of Fixations	Total Time in Zone	Face Expression(s)	Retention Test score
U1	Sc1	62	24.56	35	11.766	neutral/blank,relaxed	6%
U1	Sc2	144	62.839	31	12.177	neutral/blank,relaxed	6%
U1	Sc3	101	46.405	22	7.266	neutral/blank,relaxed	6%
U1	Sc4	111	38.126	5	2.512	neutral/blank,relaxed	6%
U2	Sc1	4	4.851	23	13.383	neutral/blank,frawn	17%
U2	Sc2	135	57.42	96	37.754	neutral/blank,frawn	17%
U2	Sc3	76	36.215	47	17.852	neutral/blank,frawn	17%
U2	Sc4	85	30.858	62	20.346	neutral/blank,frawn	17%
U3	Sc1	6	8.432	31	20.775	neutral/blank, slightly happy,relaxed	17%
U3	Sc2	59	51.21	40	26.935	neutral/blank,relaxed	17%
U3	Sc3	23	28.046	14	16.832	neutral/blank,relaxed	17%
U3	Sc4	30	23.501	25	16.33	neutral/blank,relaxed	17%

Table E.6.12: Correlated data for the non-attention-grabbing ECA (males)

	3 Females - Attention_Grabing_Humorous - Order 2 (Simple Content)							
		Background		Avatar face				
User	Scene	Number of Fixations	Total Time in Zone	Number of Fixations	Total Time in Zone	Face Expression(s)	Interruption Face Expressions	Retention Test score
U1	Sc1	143	68.624	97	40.3	blank,frawn	happy	7%
U1	Sc2	228	91.083	119	42.194	blank,frawn,boredom	slightly happy	7%
U1	Sc3	124	55.468	50	22.11	neutral/blank, surprised,frawn,tired,sad	neutral/blank,frawn	7%
U1	Sc4	124	53.881	51	19.449	skeptical, blank,tired	surprised,neutral/blank	7%
U2	Sc1	85	54.655	24	18.255	blank,relaxed,surprised	surprised,happy(slightly)	0%
U2	Sc2	156	62.218	46	18.65	blank,relaxed,attentive	neutral/blank,relaxed	0%
U2	Sc3	52	38.343	22	12.492	neutral/blank,relaxed,attentive	skeptical, neutral/blank	0%
U2	Sc4	36	27.201	4	3.302	blank,relaxed,attentive	slightly happy	0%
U3	Sc1	85	54.655	24	18.255	blank,relaxed	neutral/blank,relaxed	14%
U3	Sc2	233	90.85	156	52.889	blank,relaxed	neutral/blank,relaxed	14%
U3	Sc3	155	63.457	111	40.414	blank,relaxed	neutral/blank,relaxed	14%
U3	Sc4	129	56.789	93	34.25	blank,frawn,attentive	neutral/blank,frawn	14%

Table E.6.13: Correlated data for the attention-grabbing ECA (females)

3 Females – Order 2 (Complex Content)							
		Background		Avatar face			
User	Scene	Number of Fixations	Total Time in Zone	Number of Fixations	Total Time in Zone	Face Expression(s)	Retention Test score
U1	Sc1	65	28.447	27	11.904	neutral/blank,frawn	11%
U1	Sc2	132	66.873	50	24.037	neutral/blank,frawn	11%
U1	Sc3	89	42.717	23	8.019	neutral/blank,frawn,confusion	11%
U1	Sc4	85	30.067	45	13.561	neutral/blank,frawn	11%
U2	Sc1	0	0	0	0	neutral/blank,relaxed,skeptical,attentive	39%
U2	Sc2	42	47.091	21	19.344	neutral/blank,relaxed,attentive	39%
U2	Sc3	28	30.273	12	10.072	neutral/blank,relaxed,attentive	39%
U2	Sc4	42	24.82	21	8.651	neutral/blank,relaxed,attentive	39%
U3	Sc1	17	11.159	74	27.902	neutral/blank,relaxed	33%
U3	Sc2	116	61.077	89	40.659	neutral/blank,relaxed,attentive	33%
U3	Sc3	65	41.567	41	21.955	neutral/blank,relaxed	33%
U3	Sc4	73	36.053	53	23.128	neutral/blank,relaxed	33%

Table E.6.14: Correlated data for the non-attention-grabbing ECA (females)

Attention – Grabing	Non – Attention Grabbing
	
	
	
	

Table E.6.15: Individual Heat Map Sample (tester 6)

Attention – Grabing	Non – Attention Grabbing
	
	
	
	

Table E.6.16: Individual Heat Map Sample (tester 12)

Attention – Grabing	Non-Attention
	
	
	
	

Table E.6.17: Group Heat Maps (Group 2)

Attention – Grabing	Non – Attention Grabbing
	
	
	
	

Table E.6.18: Individual Gaze trails Sample (tester 7)

Attention – Grabing	Non – Attention Grabbing

Table E.6.19: Individual Gaze trails Sample (tester 13)